

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne démocratique et populaire  
وزارة التعليم العالي و البحث العلمي  
Ministère de l'enseignement supérieur et de la recherche scientifique  
جامعة سعد دحلب البليدة  
Université SAAD DAHLAB de BLIDA  
كلية التكنولوجيا  
Faculté de Technologie  
قسم الإلكترونيك  
Département d'Électronique



## Mémoire de Master

Filière Électronique

Spécialité Électronique des Systèmes Embarqués

présenté par

Aichaoui Fairouz

&

Ferradj Kaouther

---

# Localisation en 2D d'une source sonore basée sur l'intercorrélacion généralisée avec transformation de phase

---

Proposé par : Farid Ykhlef & Belgacem Douaer

Année Universitaire 2019-2020

## Remerciements

---

Tout d'abord, nous remercions dieu le tout puissant de nous avoir donnée le courage et la volonté et patience d'élaborer ce modeste travail.

Nous tenons à exprimer nos remerciements les plus vifs et notre gratitude la plus totale à nous encadreur Mr Farid Ykhlef et Mr Belgacem Douaer qu'ils nous orientés par leurs conseils bien utiles et leurs soutiens tout le long de notre travail.

Nous tenons à remercier les membres de jury d'avoir accepté d'examiner notre travail.

Nous sincères remerciements s'adressent à nos enseignants département d'Électronique de Blida et à tous les enseignants de notre cursus d'étude.

Enfin, nous adressons nos remerciements à toute personne ayant contribué de près ou de loin à la concrétisation de ce mémoire.

## *Dédicace*

### *Je dédie ce modeste travail*

*A mes chers parents ... aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être.*

*A mes chers frères Ahmed, Djilali, Abdelkader...*

*A toute ma famille ...*

*A mon cher adorable binôme Kaouther de m'avoir partagé les idées, les moments forts en travail, les instants de pure folie, merci de m'avoir partagé tous ces souvenirs.*

*A tous mes chers amis et collègues... plus particulièrement : Lydia*

*Vos encouragements, mais aussi pour les bons moments qui ont contribué à rendre ces années inoubliable.*

*Et à tous ceux qui m'ont aidé de près ou de loin pour l'aboutissement de ce travail.*

*Fairouz*

## *Dédicace*

### *Je dédie ce modeste travail*

*A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,*

*A mes chères sœurs Wafae, Maroua, Sarah, Yousra, Dana et Aya pour leurs encouragements permanents, et leur soutien moral, sans oublier mes adorables nièces Nline, Dania, Lina et mes deux chers neveux Adam et Idriss.*

*Une spéciale dédicace à ma chère amie et binôme Fairouz et sa famille.*

*A mes chères amies Sihem, Zola, Racha, Basma..... En souvenir de notre sincère et profonde amitié et des moments agréables que nous avons passés ensemble. Veuillez trouver dans ce travail l'expression de mon respect le plus profond et mon affection la plus sincère.*

*Et à tous ceux qui m'ont aidé de près ou de loin pour l'aboutissement de ce travail.*

*Kaouther*

---

## ملخص:

أصبح توطين مصدر الصوت لبعض الوقت محط اهتمام في العديد من المجالات والتطبيقات مثل أنظمة مؤتمرات الفيديو وأنظمة التحكم الصوتي وصوتيات الغرفة وما إلى ذلك. في هذه الأطروحة تم تطوير نظام توطين مصدر الصوت على أساس طريقة التعميم عبر الارتباط مع تحويل الطور. هنا يتم الحصول على إشارات ، ثم تُستخدم هذه الإشارات لحساب Microsoft Kinect الصوت من خلال شبكة ميكروفون متصلة بوحدة تحكم الفرق في وقت الوصول بين أزواج الميكروفون مما سيسمح بتحديد موقع ومتابعة مصدر الصوت. أظهرت النتائج التجريبية فعالية طريقة تحديد وتتبع مصادر الصوت

**كلمات المفاتيح:** توطين مصدر الصوت, طريقة تعميم الارتباط المتبادل مع تحويل الطور, الفرق في وقت الوصول.

---

**Résumé :** La localisation de source de sonore (SSL) est devenue depuis un certain temps un centre d'intérêt dans de nombreuses applications telles que les systèmes de vidéo conférence, les systèmes de commande vocale, etc..... Dans ce mémoire, un système de localisation de source sonore base sur la méthode d'intercorrélation généraliser avec transformation de phase (GCC\_PHAT) est développer. Les signaux sonores sont acquis par un réseau de microphone connecter a une console Microsoft Kinect modifié, ces signaux sont utilisés par la suite pour calculer la différence de temps d'arrives (TDOA) entre les paires de microphone qui permettra de localiser et de suivre la source sonore. Les résultats ont permis de démontrer l'efficacité de la méthode pour la localisation et le suivi de source sonore.

**Mots clés :** La localisation de source de sonore, GCC-PHAT, TDOA, l'estimation du retard temporel.

---

**Abstract:** Sound source localization (SSL) has been a focus of interest in many fields and applications such as video conferencing systems, voice control systems, room acoustics etc. In this work, a sound source localization is based on the generalized cross correlation method with phase transformation (GCC\_PHAT) to determine the direction of arrival of the sources. The signals are acquired via a microphone array connected to a Microsoft Kinect; these signals are then used to calculate the difference in arrival time (TDOA) between the pair of microphones to locate the sound source. Experimental results have demonstrated the effectiveness of the sound source localization and tracking.

**Keywords:** Sound source localization, GCC-PHAT, TDOA, time delay estimation.

---

## Listes des acronymes et abréviations

**SSL:** Sound source localisation.

**DOA:** Direction of arrival.

**TDOA:** Time difference of arrival.

**CC:** Cross Correlation.

**GCC:** Generalized Cross Correlation.

**FFT:** Fast Fourier transforms.

**IFFT:** Inverse fast Fourier transform.

**GCC-PHAT:** Generalized Cross Correlation PHase Transform.

**SRP:** Steered Reponse Power.

**SRP-PHAT:** Steered Reponse Power-PHase Transform.

**SVD:** Singular value Decomposition.

**Music:** Multiple signal classification.

**2D:** Two Dimensions.

## Table des matières

Introduction générale.....	1
----------------------------	---

### Chapitre1 : Généralité sur la localisation des sources sonores

1.1 Introduction.....	3
1.2 Notions sur l'acoustique.....	3
1.2.1 Qu'est-ce qu'un son ?.....	3
1.2.2 Emission, Propagation, Réception.....	4
1.2.3 Caractéristique d'un son.....	5
1.3 Localisation d'une source sonore chez l'être humain.....	8
1.4 Les stratégies de la localisation des sources acoustiques.....	9
1.4.1 Méthode basée sur des différences de temps d'arrivées (TDOA).....	10
1.4.2 Méthode à faible résolution (formation de voies).....	12
1.4.3 Méthode à haute résolution (approches des sous-espaces).....	13
1.5 Conclusion.....	14

### Chapitre2 : Présentation des algorithmes de localisation

2.1 Introduction.....	15
2.2 Choix de la méthode de calcul de la direction d'arrivée d'une source sonore.....	15
2.3 Les méthodes de calcul de la direction d'arrivée.....	16
2.3.1 Corrélation croisée généralisée (GCC) utilisant la transformation de phase (GCC-PHAT).....	18
2.3.2 Dérivation de la GCC.....	19
2.3.3 Formation de voies.....	22
2.3.4 Méthode à haute résolution (approches de sous-espaces).....	29
2.4 Matériels : Microsoft Kinect Xbox360.....	32

2.4.1 Définition.....	32
2.4.2 Caractéristique.....	33
2.4.3 Domaine d'application de kinect.....	36
2.5 Conclusion.....	37

### **Chapitre3 : Localisation en 2D par l'algorithme GCC-PHAT**

3.1 Introduction.....	38
3.2 Description du système.....	39
3.3 Localisation de la source (la direction d'arrivée).....	40
3.3.1 Principe.....	40
3.3.2 La résolution angulaire.....	41
3.4 La localisation en deux dimensions.....	42
3.4.1 Définition de l'hyperbole.....	43
3.4.2 Caractéristique de l'hyperbole.....	44
3.4.3 L'équation de l'hyperbole.....	44
3.4.4 Relation dans les fonctions hyperboliques.....	45
3.4.5 L'équation d'une hyperbole à partir d'un graphique.....	46
3.4.6 Tracer une hyperbole à l'aide de l'équation.....	47
3.4.7 Partie expérimentale.....	47
3.5 Conclusion.....	50

### **Chapitre4 : Résultat et Discussion**

4.1 Introduction.....	51
4.2 Expérience 1.....	52
4.3 Expérience 2.....	58
4.4 Expérience 3.....	62
4.5 Conclusion.....	65
Conclusion générale.....	66



## Liste des figures

Figure1.1 : La propagation des ondes à la surface.....	4
Figure1.2 : L'émission, la propagation et la réception du son.....	5
Figure1.3 : L'évolution de l'amplitude sonore dans le temps.....	6
Figure1.4 : Oscillation temporelle de l'intensité d'une onde sonore.....	7
Figure1.5 : Signal réel et signal échantillonné.....	7
Figure1.6 : (a) Une onde est transversale. (b) Une onde est longitudinale.....	8
Figure1.7 : Localisation de la source sonore chez l'être humain.....	9
Figure1.8 : Les stratégies de la localisation des sources acoustiques.....	10
Figure1.9 : La différence de temps d'arrivée par un réseau de deux microphones.....	11
Figure1.10 : Estimation du retard temporel par les fonctions de corrélation.....	11
Figure1.11 : Méthode de formation de voies.....	12
Figure2.1 : L'estimation de la différence de temps d'arrivées d'une onde sonore par Deux microphones.....	17
Figure2.2 : La différence de temps d'arrivées TDOA entre deux microphones.....	18
Figure2.3 : Schéma fonctionnel du GCC.....	22
Figure2.4 : Structures des formateurs de voies : (a) modèle d'un champ proche (bande étroite), (b) modèle d'un champ lointain (bande large).....	24
Figure2.5 : La Kinect XBOX360.....	33
Figure2.6 : Illustration des composants de Kinect.....	34
Figure2.7 : Réseau de microphone dans une Kinect Xbox360.....	36
Figure2.8 : Véhicule aérien sans pilote (UAV).....	38
Figure3.1 : Description globale du système.....	40
Figure3.2 : Principe de l'estimation de l'angle d'arrivées d'une source sonore.....	41

Figure3.3 : Une fonction hyperbole.....	43
Figure3.4 : L'hyperbole centrée à l'origine.....	44
Figure3.5 : L'hyperbole a l'axe horizontal et vertical.....	46
Figure3.6 : Illustration de la configuration de la simulation.....	48
Figure4.1 : Les signaux reçus par les 4 microphones.....	51
Figure4.2 : plot d'intercorrélation (xcoor).....	53
Figure4.3 : plot d'intercorrélation généralisée (GCC-PHAT).....	53
Figure4.4 : Exemple d'expérimentation l'expérience 1.....	54
Figure4.5: L'estimation de la direction d'arrivées sous MATLAB.....	54
Figure4.6 : Comparaison des erreurs entres les valeurs D.....	57
Figure4.7 : Exemple d'expérimentation l'expérience 2.....	58
Figure4.8 : La tendance des deux hyperboles.....	62
Figure4.9 : Exemple d'expérimentation l'expérience 3.....	63
Figure4.10 : La localisation de la source sonore en deux dimensions.....	64

## Liste des tableaux

Tableau 2.1 : Les fonctions de pondération utilisées dans la GCC.....	21
Tableau 3.1 : Relation entre hyperbole de base et transformée.....	45
Tableau 4.1 : Caractéristiques et informations des éléments utilisées.....	51
Tableau 4.2 : L'estimation de l'angle d'arrivée avec la variation de la distance entre les microphones.....	55
Tableau 4.3 : L'estimation de l'angle d'arrivée avec la variation de $F_s$ .....	59
Tableau 4.4 : Les coordonnées proposée pour la source sonore.....	64

# Introduction générale

---

Dans des environnements acoustiques, l'information sur la position des sources joue un rôle très important dans des applications telles que la poursuite automatique du locuteur par une caméra pour la vidéoconférence, navigation robotisée, etc. L'estimation des positions des sources, souvent appelé localisation de sources a eu un considérable intérêt dans la recherche pendant des décennies.

Un réseau de plusieurs microphones est utilisé pour estimer les angles (direction) d'arrivées des sources acoustiques. Pour l'estimation du nombre des sources et leurs localisations, plusieurs algorithmes intéressants existent pour la localisation surtout dans le cas d'une source sonore assez éloignée et résoudre les problèmes liés à cette opération.

La localisation d'une source acoustique est un procédé bien développé du système auditif humain. Utilisant seulement deux capteurs, ce système biologique a une remarquable précision en résolvant la position des locuteurs et d'autres sources acoustiques. Avec deux oreilles et un traitement fait par le cerveau, l'être humain a la capacité de résoudre tous les problèmes de l'environnement.

Dans ce mémoire, l'idée traitée consiste à employer un réseau de quatre microphones en utilisant le dispositif Kinect Xbox 360 de Microsoft. Ensuite exploiter les signaux en temps réel pour : premièrement, estimer la direction d'arrivée d'une source sonore par un ensemble de deux microphones. La deuxième étape consiste à localiser les coordonnées de la source par un modèle en deux dimensions.

L'acquisition des signaux et leur prétraitement est réalisée en temps réel grâce au dispositif de capteurs audio (Microsoft Kinect Xbox 360) connectée à un ordinateur portable via un port USB. Ceci est pour explorer les possibilités de ce moyen beaucoup

moins coûteux que d'autres systèmes d'acquisition de données dans le marché.

Ce mémoire se compose de trois chapitres et ils sont organisés comme suit :

- Le premier chapitre de ce mémoire est dédié à la localisation des sources sonores et ses principes, les notions de base de la localisation acoustique, les stratégies de la localisation existantes sont traitées globalement en détaillant la méthode utilisée à la fin du chapitre.
- Dans le deuxième chapitre, le principe et les structures des méthodes de la localisation sont présentés.
- Les expériences réalisées avec les résultats obtenus suivis des interprétations sont présentées dans le troisième chapitre.
- À la fin de ce mémoire, une synthèse des résultats obtenus est présentée dans la conclusion générale.

# Chapitre 1 Généralités sur la localisation des sources sonores

---

## 1.1 Introduction

Localiser une onde sonore revient à déterminer sa direction d'arrivée ainsi que sa distance en utilisant un réseau de microphone qui consiste de deux ou plusieurs microphones positionnés de telle manière à profiter au maximum de l'information spatiale, les oreilles humaines avec le cerveau peuvent exactement localiser et poursuivre les sources. Dans la littérature, il existe plusieurs méthodes pour la détermination de la direction ou bien localiser sa coordonné. La localisation d'une source audio et la poursuite automatique dans un environnement réverbérant sont des tâches provocantes et souvent nécessaires dans des applications audio telles que le téléphone mobile à mains libre, la reconnaissance de la parole, etc.

Dans ce chapitre, Nous allons présenter premièrement dans ce chapitre une introduction sur les notions de base de la localisation des sources acoustiques. Une introduction sur Les diverses algorithmes du traitement de signal dédiés à la localisation des sources audio sont présentés. Ces algorithmes peuvent être utilisés pour l'estimation de l'angle d'arrivée de l'onde sonore traversant un réseau de microphone.

## 1.2 Notions sur l'acoustique

### 1.2.1 Qu'est- ce qu'un son ?

Le son est une onde produite par la vibration mécanique d'un support fluide ou

solide et propagée grâce à l'élasticité du milieu environnant sous forme d'ondes longitudinales, par extension physiologique, le son désigne la sensation auditive à laquelle cette vibration est susceptible de donner naissance.

Lorsqu'on jette une pierre dans l'eau, on peut facilement observer le phénomène de propagation des ondes à la surface :



**Figure 1.1** : la propagation des ondes à la surface.

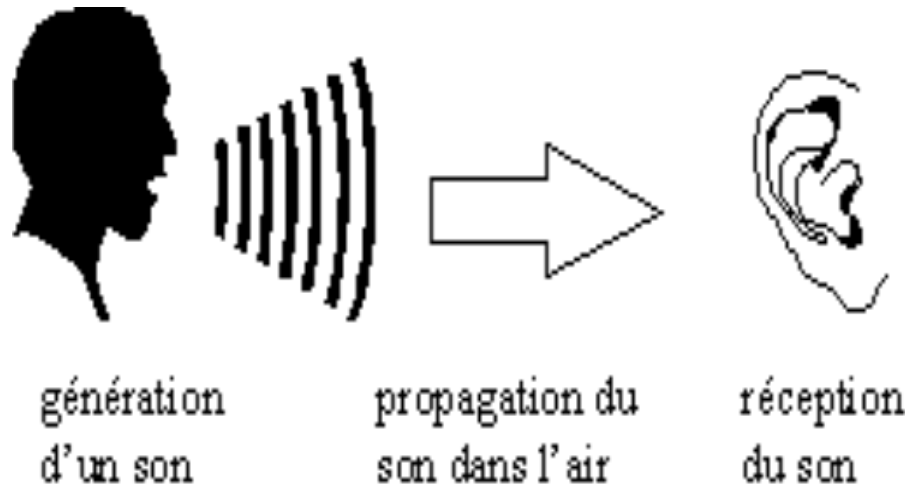
Lors de la diffusion d'un son dans un concert, c'est l'air qui permet sa transmission jusque nos oreilles. De même que l'exemple de l'eau illustrée ci-dessus, les molécules d'air transmettent l'énergie et son donc un support pour le son.

### **1.2.2 Emission, propagation, réception**

Pour qu'un soit émis, une énergie doit avant tout mettre en mouvement un corps pour produire une vibration. Ainsi, le muscle du larynx, la chute d'un objet sur le sol, ou la tension électrique dans un haut-parleur, provoqueront l'énergie nécessaire pour produire cette vibration. Ensuite, pour que ce son puisse se propager, il faut un milieu pour la transmission de la vibration. En créant des surpressions ou des dépressions, l'air permet la propagation de l'onde. Les matériaux solides ont aussi cette capacité de transmettre le son. Dans le vide par contre, aucun son ne peut se propager, car il n'y a aucun de support.

Enfin, pour être perçue, il doit y avoir un récepteur sensible. Chez l'homme, l'oreille possède une membrane (voir figure 1.2) capable de transmettre les informations de vibration en signaux nerveux jusqu'au cerveau, grâce au nerf auditif. De même, le

microphone possède également une membrane permettant de transformer le déplacement de l'air en signaux électrique.



**Figure 1.2** : l'émission, la propagation et la réception du son.

Les ondes sonores se déplacent à environ 344 m/s (mètres par seconde) dans de l'air à une température de 20°C (Celsius), la vitesse qu'on peut arrondir à environ un kilomètre toutes les trois secondes, ce qui est utile pour mesurer grossièrement la distance d'un éclair lors d'un orage, la vitesse de la lumière rendant sa perception quasi instantanée. En milieu non gazeux, le son peut se propager encore plus rapidement, dans l'eau, sa vitesse est de 1482 m/s ; dans l'acier de 5050 m/s. Il ne se propage pas dans le vide car il n'y a pas de moyen pour supporter les ondes produites. Le vide est donc un excellent isolant phonique.

### **1.2.3 Caractéristique d'un son**

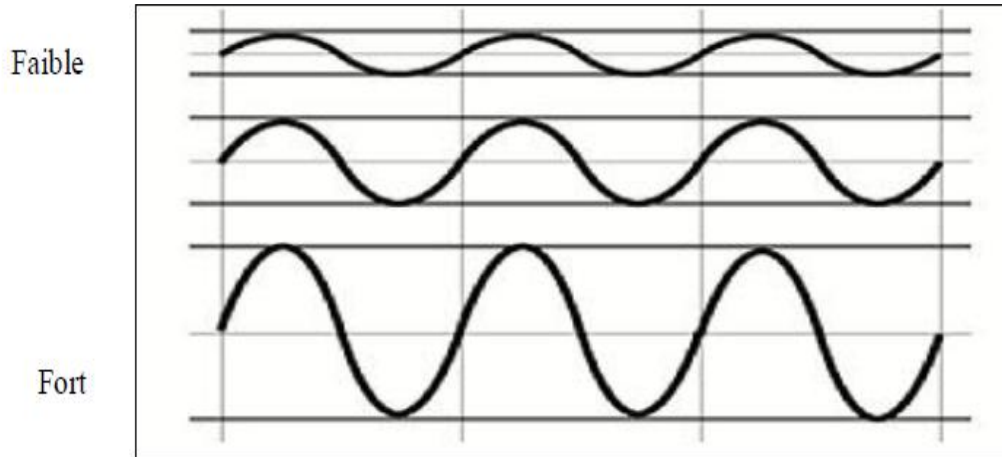
Comme tout phénomène vibratoire, le son peut être analysé comme un signal qui varie dans le temps. Les caractéristiques essentielles sont l'amplitude et la fréquence

#### ***a. L'amplitude***

La première caractéristique d'un son est l'amplitude. Appelée aussi intensité ou volume sonore, c'est l'expression de la pression de l'air qui se mesure en décibels



(dB). Pour 0 dB, il correspond au niveau de pression acoustique minimal pour qu'un son puisse être perçu de nos oreilles.



**Figure 1.3** : L'évolution de l'amplitude sonore dans le temps.

Exemple concret :

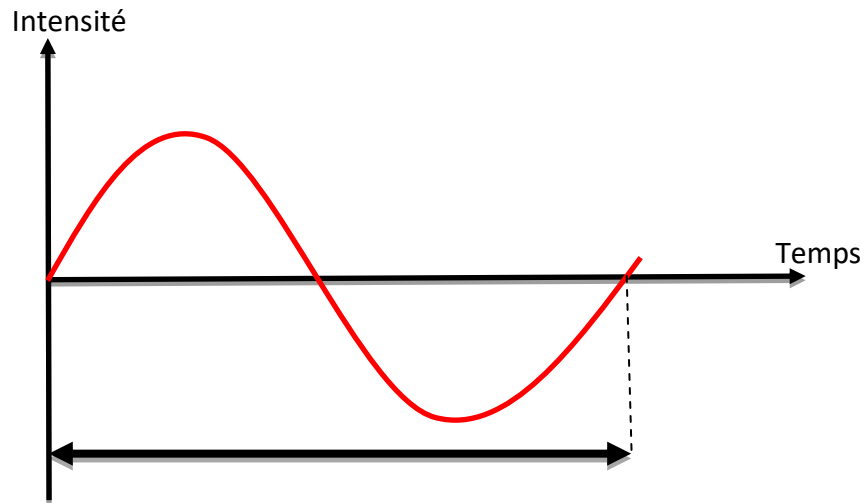
- De 0 à 10 dB : Seuil d'audibilité, Désert
- De 30 à 40 dB : forêt
- De 40 à 50 : conversation
- De 60 à 70 dB : sonnerie de téléphone
- De 80 à 90 dB : tondeuse à gazon, klaxon de voiture
- 120 dB : seuil de la douleur, décollage d'un avion
- 180 dB : décollage de la fusée Ariane, lancement d'une roquette

### ***b. Fréquence***

La fréquence est le **nombre de périodes par unité de temps** ce qui correspond à l'inverse de la période :  $f = 1/T$  ou  $f$  est la fréquence en Hertz (Hz ou  $s^{-1}$ ) et  $T$  la période en seconde (s).

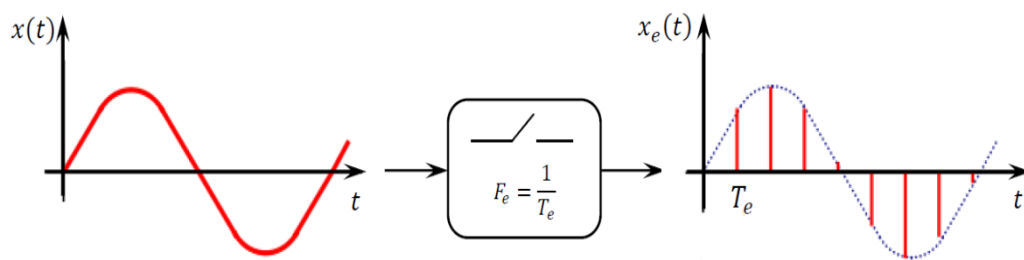
La fréquence, exprimée en Hertz (Hz), est le nombre de répétition d'une période par seconde. Plus elle est élevée et plus le son paraîtra « aiguë », à l'inverse, il paraîtra « grave ». En musique, la fréquence définit donc la hauteur d'un son, soit, la note.

(Exemples : la note « LA » correspond à 440Hz, soit 440 vibration en une seconde)[1].



**Figure 1.4 :** Oscillation temporelle de l'intensité d'une onde sonore.

L'échantillonnage est une opération qui consiste à transformer un signal analogique en un signal numérique (voir Figure 1.5). Cette opération est fondamentale en traitement de signal car à l'issue de l'opération le signal obtenu étant numérique il est possible de le traiter au moyen d'un ordinateur.



**Figure 1.5:** Signal réel et signal échantillonné.

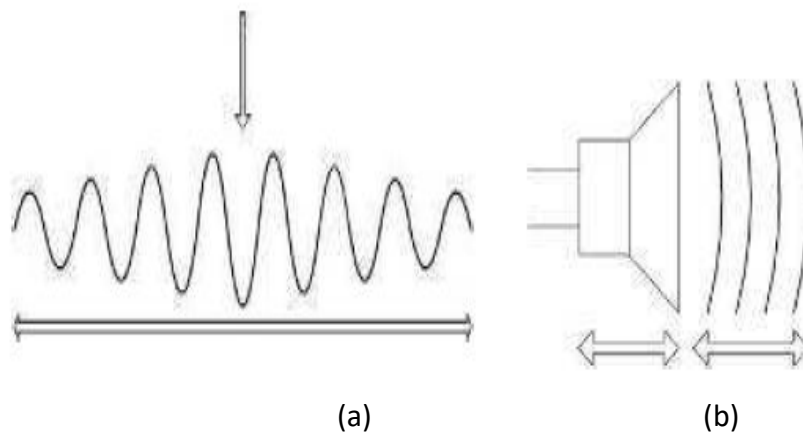
Le principe d'échantillonnage concerne la transformation d'un signal à temps continu en un signal à temps discret. Dans cette opération, une partie de l'information est habituellement perdue et il convient de bien comprendre le phénomène d'échantillonnage pour connaître et si possible minimiser l'information perdue [2].

### **c. Longueur d'onde**

La longueur d'onde est la distance séparant deux molécules successives dans le même état vibratoire (même pression et vitesse acoustique) ou encore la distance parcourue par l'onde pendant une période.

### **d. Onde transversale et longitudinale**

Une onde est transversale lorsque les particules oscillent dans une direction perpendiculaire la direction de propagation de l'onde (voir Figure 1.6 (a)). Une onde est longitudinale lorsque les particules vibrent dans la même direction la propagation de l'onde (voir Figure 1.6 (b)) [3].



**Figure 1.6:** (a) Une onde est transversale. (b) Une onde est longitudinale.

## **1.3 Localisation d'une source sonore chez l'être humain**

A nos yeux l'audition binaurale c'est une condition très nécessaire et indispensable pour localiser la provenance d'un son et sa distance, mais également de se focaliser sur une source utile dans un milieu bruyant.

Scientifiquement les ondes sonores arrivent sur le tympan qu'elles font vibrer. Ces vibrations sont transmises, via les osselets, à la fenêtre ovale. Après la fenêtre ovale, les ondes sont transmises à l'endolymphe et à la membrane basilaire de la cochlée, sur laquelle se trouvent des cellules ciliées.

D'une façon globale l'oreille de l'être humain fonctionne par plusieurs indices pour localiser une source sonore dans un espace, comme ceci :

Sur le plan horizontal résulte des différences signaux parvenant aux deux oreilles par l'utilisation des principaux indices, Dans le plan vertical, les principaux indices utilisés sont des indices monauraux, ils proviennent de la modification d'un son par le torse, la tête et l'oreille externe d'un auditeur. La différence de temps d'arrivée d'une onde sonore à chaque oreille est un indice important pour estimer la position d'une source dans le plan horizontal [4].

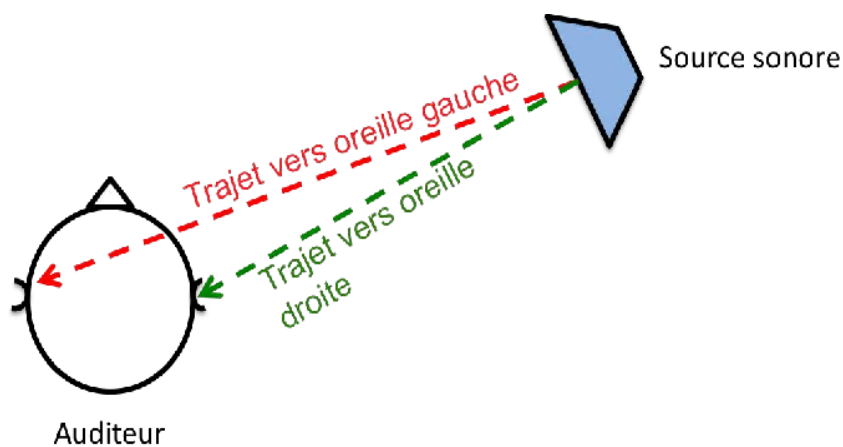
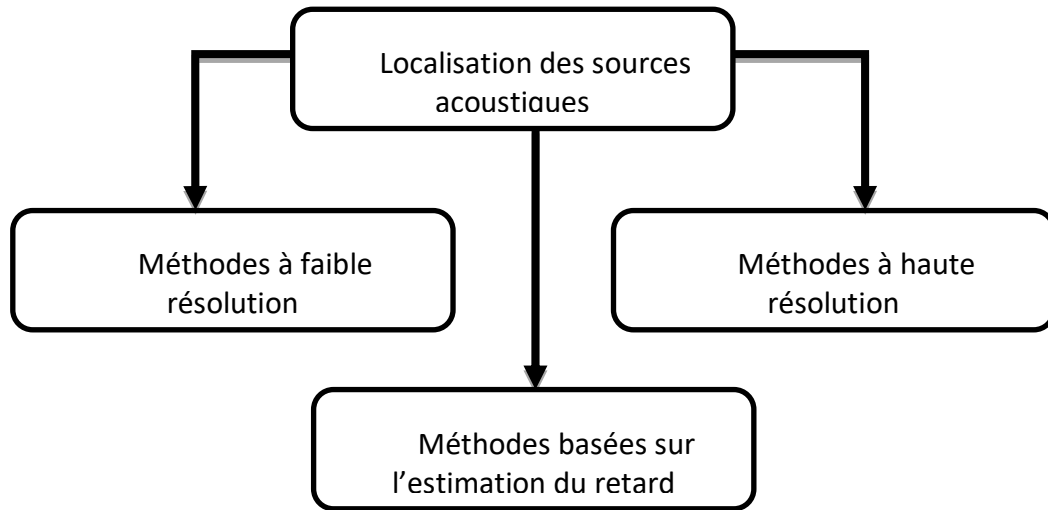


Figure 1.7 : Localisation de la source sonore chez être humain.

## 1.4 Les stratégies de la localisation des sources acoustiques

Les procédures de localisation des sources peuvent être classifiées en trois grandes catégories (voir la figure 1.8) : celles basées sur la maximisation de la réponse d'un formateur de voies (SRP : Steered Response Power), techniques adoptant les concepts d'estimation à haute résolution spectrale et les approches exploitant le retard temporel (TDOA). Ces grandes classes diffèrent selon l'environnement d'application et la méthode d'estimation. Dans la première classe, l'estimé de la position est obtenu directement après le formateur de voies retard et somme des données reçues au niveau des capteurs. La deuxième est basée sur le

calcul de la matrice de corrélation et la dernière catégorie cherche à transformer le retard temporel entre les éléments du réseau en positions.

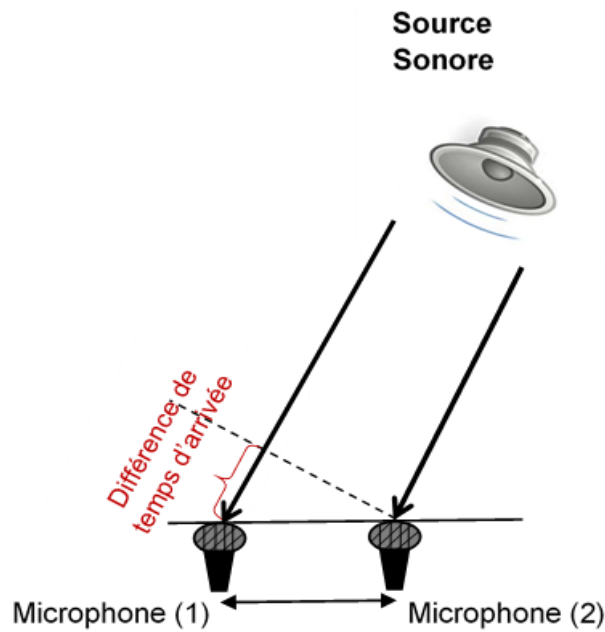


**Figure1.8** : Les stratégies de la localisation des sources acoustiques.

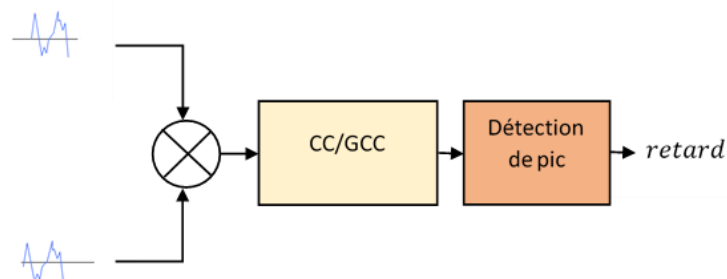
#### **1.4.1 Méthode basée sur des différences de temps d'arrivée (TDOA)**

La première catégorie est basée sur le calcul la différence du temps (TDOA) entre les signaux reçus qui nous permet de calculer l'emplacement de la source. L'une des méthodes les plus utilisées pour déterminer le retard entre tes signaux : La fonction d'intercorrélation (CC) et la fonction d'intercorrélation généralisée (Generalized Cross Correlation GCC). Ces localisateurs utilisent les données de délai de chaque paire de microphones (angle d'arrivée) ainsi que les emplacements connus des microphones. L'estimation de la direction assez compliquée par la présence de bruit et de réverbérations dans les capteurs.

Dans de nombreuses applications du traitement du signal il est nécessaire de comparer des signaux entre eux (voir Figure 1.9), en particulier pour déterminer un degré de ressemblance ou pour synchroniser des horloges. À partir de ce principe, la méthode d'intercorrélation est utilisée pour estimer le TDOA. Cette méthode permet aussi de déterminer la direction de la source sonore par l'estimation de la différence de temps.



**Figure 1.9 :** la différence de temps d'arrivée par un réseau de deux microphones.



**Figure 1.10 :** Estimation du retard temporel par les fonctions de corrélation.

Les méthodes basées sur le TDOA sont moins coûteuses en termes de calcul, mais elles ont des limites car elles supposent un modèle à source unique. La présence de plusieurs sources sonores simultanées (ce qui est souvent le cas dans les applications de localisation de sources sonores) tel qu'un bruit ambiant ou des niveaux de réverbération modérés, entraînent une fiable estimation.

Dans une première étape un ou plusieurs délais entre différentes paires de microphones (c.-à-d. le TDOAs) sont estimés. En utilisant ces évaluations de TDOA

est la géométrie du réseau de microphone, la position des sources peut être calculée dans une deuxième étape. Cette catégorie inclut les méthodes de l'intercorrélation conceptuellement simples qui sont largement utilisées [5][6][7].

#### 1.4.2 Méthode à faible résolution (formation de voies)

La formation de voies est parmi les premières méthodes développées pour la localisation des sources rayonnantes. Elle est basée sur le calcul de l'énergie reçue dans une direction privilégiée en réalisant des rotations virtuelles de l'antenne. Cette méthode est dite à faible résolution car sa résolution spatiale dépend de la largeur du diagramme de réception d'un signal (onde sonore).

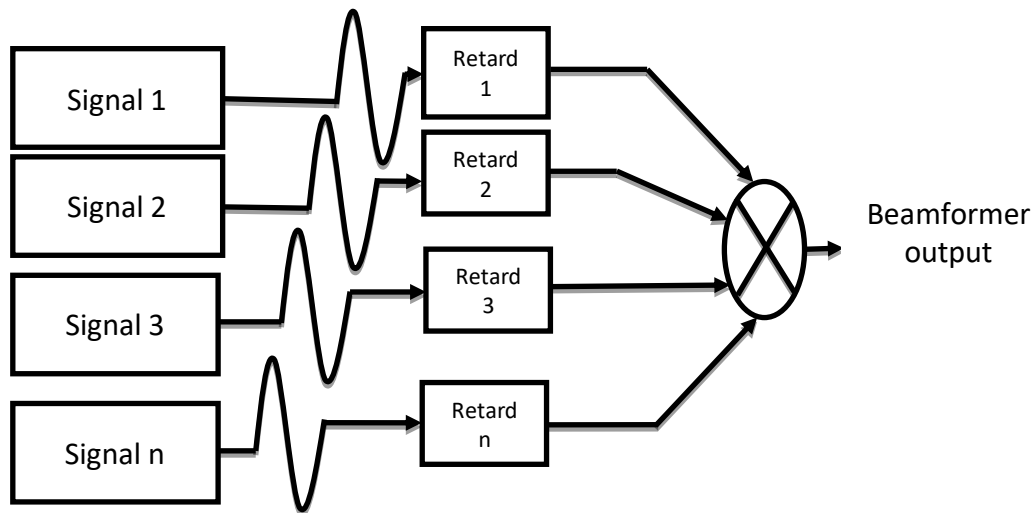


Figure1.11 : Méthode de formation de voies.

La formation de voie est une méthode permettant la focalisation des ondes dans une direction spécifique de l'espace. Typiquement, une formation de voie exploite les signaux de N microphones et en les combinant linéairement par l'intermédiaire de filtres de réponse impulsionnelle de façon à obtenir un signal d'une seule sortie.

Ce type de formateur consiste à appliquer des retards sur les signaux reçus de réseaux en vue de compenser le retard de la propagation. Les signaux alignés temporellement et sommés forment un seul signal à la sortie. Cette méthode est utilisée dans le cas d'un réseau de plus de deux microphones [8][9].

### 1.4.3 Méthode à haute résolution (approches des sous-espaces)

Les méthodes dites à Haute Résolution (HR), qui trouvent leurs applications en traitement d'antenne comme en analyse spectrale [Marcos 1998], présentent l'avantage de s'affranchir des limitations naturelles de l'analyse de Fourier. En effet, en l'absence de bruit, leur précision et leur résolution fréquentielles sont virtuellement infinies (bien qu'en pratique limitées par la précision finie des machines de calcul). Ceci est rendu possible en s'appuyant fortement sur un modèle paramétrique de signal pour la localisation des sources acoustiques. Ainsi, contrairement à l'analyse de Fourier qui consiste à représenter le signal dans un domaine transformé, les méthodes HR sont des méthodes d'estimation paramétrique. Dans le cadre du traitement du signal audio, elles restent peu utilisées en raison de leur forte complexité algorithmique.

L'origine des méthodes HR remonte aux travaux de Prony publiés en 1795, qui visent à estimer une somme d'exponentielles par des techniques de prédiction linéaire [Riche de Prony, 1795]. Plus récemment, cette approche a été approfondie les signaux sinusoïdes estimés. Les méthodes hautes résolution ont été développées à la fin des années 1970. Différentes méthodes existent, toutes basées sur le principe d'une décomposition de la matrice inter-spectrale en deux sous-espaces, le sous-espace source et le sous-espace bruit (ou sous-espace orthogonal).

Ces méthodes exploitent les statistiques d'ordre deux des signaux enregistrés. L'idée de base de ces méthodes est la décomposition de l'espace d'observation en deux sous espaces complémentaires : le sous-espace signal et le sous-espace bruit. C'est Pisarenko qui a eu l'idée en 1973 de faire une Décomposition en Valeurs Singulières (SVD) (singular value décomposition) sur la matrice d'autocorrélation du signal, et d'exploiter les données, en montrant que les plus petites valeurs propres correspondaient au bruit, ce qui permettait de faire une décomposition de l'espace et de déduire les angles en projetant une famille de vecteurs paramétrés par leur direction d'arrivée. Parmi les méthodes employées les plus extraites de ce principe de SVD est la méthode de "Multiple Signal Caractérisation" (MUSIC).

Les méthodes HR modernes reposent sur les propriétés particulières de la matrice



de covariance du signal. Ainsi, l'étude de son rang permet de séparer l'espace des données en deux sous-espaces, l'espace signal engendré par les sinusoides, et l'espace bruit qui est son complémentaire orthogonal. Les méthodes HR issues de cette décomposition en sous-espaces sont plus robustes que les techniques de prédiction linéaire. Citons par exemple la méthode Multiple Signal Classification (MUSIC) [Schmidt, 1986], qui s'appuie sur l'espace bruit, et les méthodes Matrix Pencil [Hua et Sarkar, 1990] et Estimation of Signal Paramètres via Rotational Invariance Techniques (ESPRIT) [Roy et al. 1986], qui s'appuient sur l'espace signal [10] [11].

## **1.5 Conclusion**

Ce chapitre d'introduction a été pour nous l'occasion d'introduire l'ensemble des notations et notions nécessaires pour la suite de ce manuscrit. Nous y avons tout d'abord donné des notions sur l'acoustique et une généralité historique sur la localisation d'une source sonore (son, microphone) aussi chez l'être humain et le principe de quelques méthodes de localisation d'une source.

Les différentes stratégies de localisation sont décrites. Les méthodes basées sur des différences de temps d'arrivées TDOA semblent très efficaces dans le cas où les sources soit sont très proches ou bien une onde sonore assez éloignée, les méthodes de TDOA ne nécessitent pas un grand nombre d'éléments de réseau, et l'utilisation des plusieurs microphones pourrait nous a déterminer l'emplacement de la source avec une grande précision résolution. La résolution dans les méthodes basées sur l'estimation du retard temporel dépend de la fréquence d'échantillonnage et la géométrie du réseau, ces méthodes sont moins complexes en termes de temps de calcul, donc beaucoup adéquats pour des applications en temps réel.

# Chapitre 2 Présentation des algorithmes de localisation

---

## 2.1 Introduction

Le domaine de la localisation d'une source acoustique est un domaine de recherche très actif et présent dans de nombreux champs d'applications par les traitements de signaux (audio surveillance, téléconférence, domaine de la défense, etc...). L'objectif des techniques de localisation d'une source sonore est de déterminer la direction de la source aussi que sa distance par rapport au réseau de microphones qui nous permet de compter et combiner l'information audio.

Nous présentons dans ce chapitre les méthodes de base les plus utilisées dans le domaine de la localisation acoustique. La classification que nous adoptons pour ces méthodes et qui sont brièvement présentées au sein du chapitre 1 repose sur le nombre de microphones disponibles dans le réseau installé et le principe de calculer le retard entre les signaux reçus. Nous présentons en premier le principe de fonctionnement et la dérivation mathématique la méthode basée sur l'estimation du retard temporel avec les capteurs utilisés et ensuite nous présentons l'appareil et la disposition du réseau utilisé dans la partie pratique.

## 2.2 Choix de la méthode de calcul de la direction d'arrivée d'une source sonore

Pour réaliser notre travail nous allons opter pour la méthode TDOA. L'une des techniques les plus utilisées est les techniques de la différence du temps (TDOA) qui

permet d'obtenir de bons résultats de localisation. Si l'on considère le cas d'un réseau de deux capteurs (microphones), le principe de l'estimation de la direction d'arrivée d'une onde sonore repose sur le calcul de ses temps de capture par chaque microphone puis en déduire le retard ( $\tau$ ) ou la différence des temps d'arrivées (TDOA) enregistrée entre ces deux microphone.

Cette méthode est beaucoup plus précise et robuste dans le cas d'une source sonore assez éloignée (cas des ondes planes) ce qui correspond aux objectifs que nous avons fixés. C'est pour cette raison que la méthode TDOA a été retenue le reste de nos calculs.

## 2.3 Les méthodes de calcul de la direction d'arrivée

On considère un réseau de deux microphones installé dans plan horizontal et une distance entre eux. Le signal de premier microphone (microphone A) est donné comme suit :

$$x_A(t) = s(t) * h_A(t) + n_A(t) \quad (2.1)$$

Et le signal à l'autre microphone (B) :

$$x_B(t) = s(t - \tau_{AB}) * h_B(t) + n_B(t) \quad (2.2)$$

Notez que pour être précis, il faudrait inclure le délai  $\tau_{AB}$  dans le signal source  $s(t)$ , c'est-à-dire  $s(t - \tau_{AB})$  dans l'équation (2.2) pour montrer que le signal reçu au microphone B est une version retardée du signal source. Ici, le problème est la différence de temps relative d'arrivée  $\tau_{AB}$ , entre ces deux microphones A et B.

Afin de déterminer l'angle  $\alpha$ , nous appliquons les formules de la trigonométrie pour un triangle de ( $\Delta d = d_A - d_B = \tau_{AB} * c$ ) et d'hypoténuse (D) :

$$\cos(\alpha) = \left( \frac{\tau_{AB} * c}{D} \right) \quad (2.3)$$

où :

$$\alpha = \cos^{-1} \left( \frac{\tau_{AB} * c}{D} \right) \quad (2.4)$$

Avec :

$d_A$  : La distance entre la source et microphone(A).

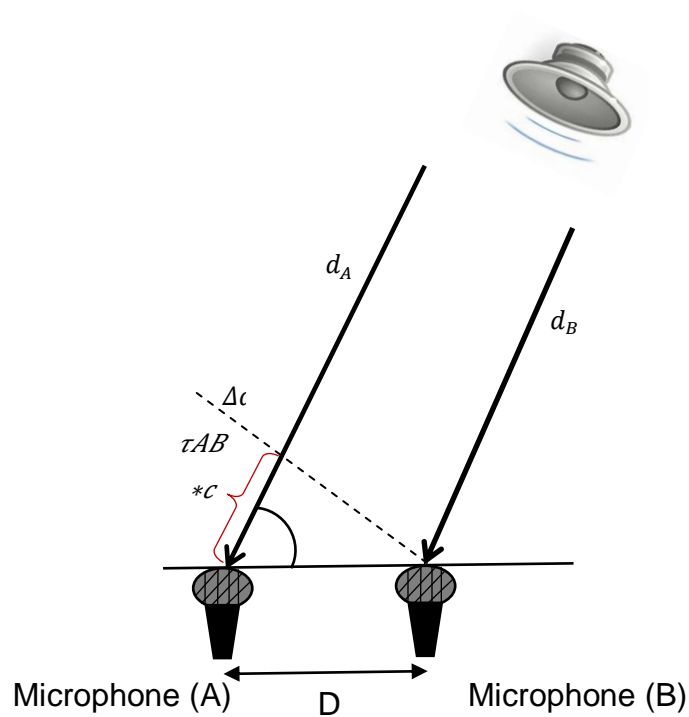
$d_B$  : La distance entre la source et microphone(B).

D : la distance entre le microphone(A) et microphone(B).

C : vitesse de propagation du son dans l'air (340m/s).

$\tau_{AB}$  : La différence de temps d'arrivée (TDOA) entre les deux microphones.

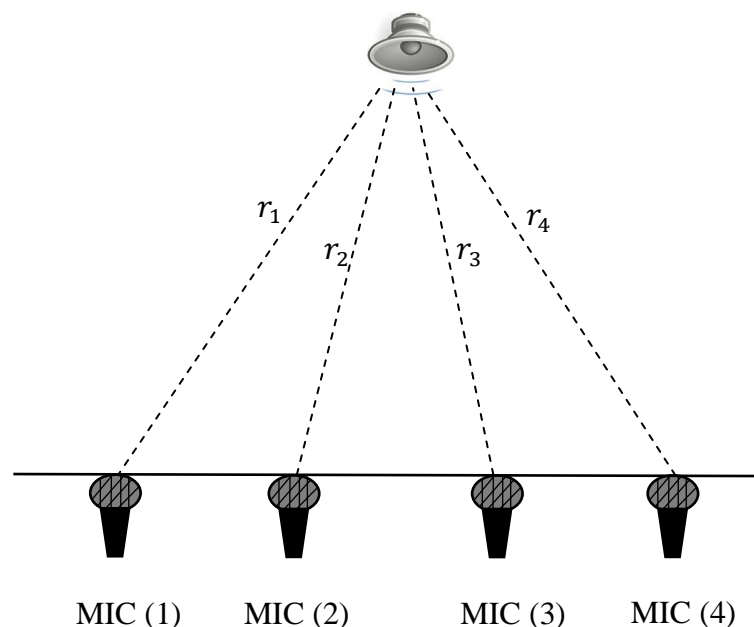
$\alpha$  : L'angle qui représente la direction d'arrivée de la source sonore.



**Figure 2.1** : Estimation de la différence de temps d'arrivée d'une onde sonore par deux microphones.

### 2.3.1 Corrélation croisée généralisée (GCC) utilisant la transformation de phase (GCC-PHAT)

Dans la littérature, il existe plusieurs méthodes pour la détermination de la direction d'arrivée d'une onde sonore. L'une des techniques les plus utilisées est la technique de la différence du temps d'arrivée (TDOA) qui permet d'obtenir de bons résultats de Localisation. Si l'on considère le cas d'un réseau de deux capteurs (microphones), le principe de l'estimation de la direction d'arrivée d'une onde sonore repose sur le calcul de ses temps de capture par chaque microphone puis en déduire le retard ( $\tau$ ) ou la différence des temps d'arrivée (TDOA) enregistrée entre ces deux microphones. Ainsi, à l'instar du système auditif humain, l'exploitation du retard que met un son pour arriver à deux capteurs (microphones) pourrait nous conduire à déterminer la direction d'arrivée de ce son. La GCC une méthode populaire pour déterminer la différence de temps d'arrivée TDOA entre deux microphones à partir de plusieurs valeurs de TDOA, on peut estimer l'emplacement de la source. Prenez un exemple d'un réseau de 4 microphones comme :



**Figure 2.2 :** La différence de temps d'arrivée TDOA entre deux microphone.

Si la distance entre les microphones  $r_i$  ( $i = 1, 2, 3, 4$ ), le temps d'arrivée (traveling time) à chaque microphone du signal de la source a ce microphone est :

$$\tau_i = r_i / c \quad (2.5)$$

La différence de temps d'arrivée (TDOA) entre deux microphone  $A$  et  $B$  peut être défini comme :

$$\tau_{AB} = \tau_A - \tau_B = (r_A - r_B) / c \quad (2.6)$$

à partir de cette relation entre le TDOA et les distances entre la source et les microphones  $r_i$ , on peut estimer l'endroit de la source sonore à partir de plusieurs TDOA entre chaque paire de microphone en utilisant plusieurs techniques, telles que : la méthode de l'intercorrélation généralisée (GCC) qui a été proposée par Knapp et Carter en 1976. Nous savons maintenant comment estimer l'emplacement de la source à partir des TDOA. La question suivante est de savoir ce qu'est la GCC et comment définir la TDOA à partir de GCC.

### 2.3.2 Dérivation de la GCC

Pour simplifier les choses, nous avons normalisé les équations ici de sorte que le retard de la source au microphone  $A$ ,  $\tau_A$  soit de zéro. En d'autres termes, nous ne nous intéressons qu'à la différence de temps d'arrivée  $\tau_{AB}$  entre les deux microphones  $A$  et  $B$ .

La fonction de l'intercorrélation généralisée GCC de ces deux signaux de microphones montrera un pic au décalage temporel où ces deux signaux décalés sont alignés, correspondant au TDOA,  $\tau_{AB}$ .

L'intercorrélation de  $x_A(t)$  et  $x_B(t)$  est définie comme suit :

$$C_{AB}(\tau_{AB}) = \int_{-\infty}^{+\infty} x_A(t) x_B(t - \tau_{AB}) dt \quad (2.7)$$

Par l'utilisation de l'équation de l'intercorrélation (2.7), La TDOA entre deux microphones  $A$  et  $B$  est le décalage  $\tau_{AB}$  qui maximise la GCC dans les signaux de données limitées par la distance entre les microphones.

Une fois que le maximum de l'intercorrélation entre  $s_1$  et  $s_2$  est déterminé (retard  $\tau$  entre ces deux signaux), nous pouvons déduire l'endroit de ce maximum en appliquant la fonction d'argument du maximum:

$$\tau_{AB} = \text{Arg\_max} (C_{AB}) \quad (2.8)$$

On peut déterminer la fonction de GCC à partir de la CC. Pour cela, en prenant la transformée de Fourier FFT de la fonction de corrélation, on obtient une (Cross Power Spectrum)

$$C_{AB}(w) = \int_{-\infty}^{+\infty} C_{AB}(\tau_{AB}) e^{j\omega\tau_{AB}} d\tau \quad (2.9)$$

En appliquant les propriétés de convolution de la transformée de Fourier pour (2.7) on (2.9) nous avons :

$$C_{AB}(w) = X_A(w) X_B^*(w) \quad (2.10)$$

où  $X_i(w)$  est la transformée de Fourier du signal  $x_i(t)$  et '\*' désigne le conjugué complexe. La transformée de Fourier inverse IFFT de (2.10), nous donne la fonction de l'Intercorrélation généralisée en terme de FFT du signal aux microphones.

$$C_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_A(w) X_B^*(w) e^{j\omega\tau_{AB}} d\omega \quad (2.11)$$

L'Intercorrélation généralisée (GCC) de  $x_A(t)$  et  $x_B(t)$  est l'Intercorrélation généralisée de leurs versions filtrées. Désignant la TF de ces deux filtres  $\psi_A(\omega)$  et  $\psi_B(\omega)$ , nous avons la GCC  $R_{AB}(\tau)$  :

$$R_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\psi_A(\omega) X_A(w)] [\psi_B(\omega) X_B^*(w)] e^{j\omega\tau_{AB}} d\omega \quad (2.12)$$

Nous définissons une fonction de pondération combinée  $\psi_{AB}(\omega)$  comme :

$$\psi_{AB}(\omega) \equiv X_A(\omega) X_B^*(\omega) \quad (2.13)$$

Les fonctions de pondération les plus utilisées avec la GCC sont résumées dans le tableau.

Approche	Fonction de pondération
Intercorrélation classique (CC)	$\psi(w) = 1$
Transformation de phase (PHAT)	$\psi(w) = \frac{1}{ X_A(\omega) X_B^*(\omega) }$
Roth (Wiener-Hopf weighting)	$\psi(w) = \frac{1}{X_A(\omega) X_A^*(\omega)}$ ou $\frac{1}{X_B(\omega) X_B^*(\omega)}$
Smoothed Coherence Transform (SCOT)	$\psi(w) = \frac{1}{X_A(\omega) X_A^*(\omega)} \cdot \frac{1}{X_B(\omega) X_B^*(\omega)}$

**Tableau 2.1** : Les fonctions de pondération utilisées dans la GCC.

Le type de pondération utilisé avec GCC est essentiel aux performances de localisation. Parmi plusieurs types de pondération, la transformation de phase (PHAT) est le pré filtre le plus couramment utilisé pour le GCC parce qu'il est robuste contre la réverbération. Le GCC avec la transformation de phase (GCC-PHAT) est performant dans un environnement réverbérant doux. Malheureusement, en présence de niveaux de réverbération même modérés, l'algorithme est sérieusement entravé en raison de la présence de pics parasites. En remplaçant (2.13) sur (2.10), la GCC devient :

$$R_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi_{AB}(\omega) X_A(\omega) X_B^*(\omega) e^{j\omega\tau} d\omega \quad (2.14)$$

La GCC-PHAT donc est donné par :

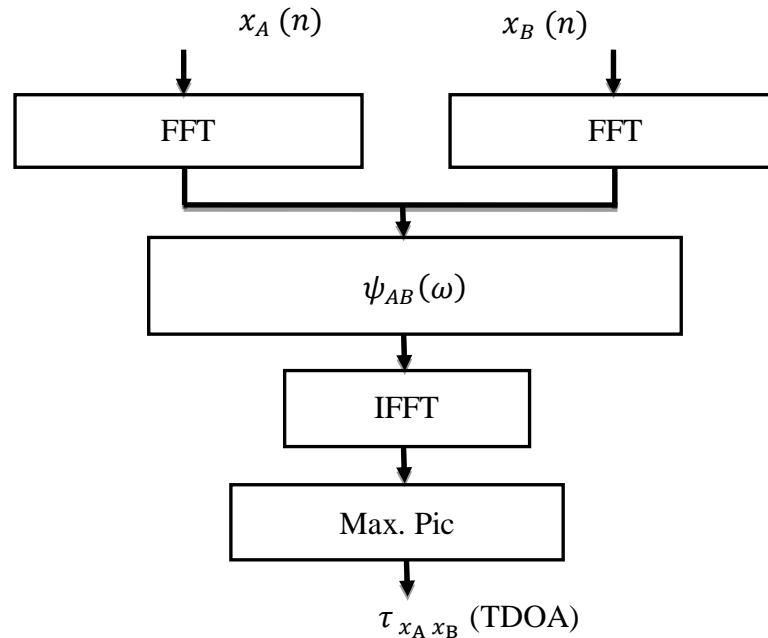
$$R_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{|X_A(\omega) X_B^*(\omega)|} X_A(\omega) X_B^*(\omega) e^{j\omega\tau} d\omega \quad (2.15)$$

Et par la même méthode et à partir de l'équation (2.14), on peut appliquer l'équation (2.8) pour estimer la TDOA.

La différence entre estimer le retard temporel par le calcul direct de l'intercorrélacion par la formule (2.7) ou l'estimer via la transformée de Fourier inverse de la densité inter-spectrale par la formule (2.15) réside dans la complexité en terme de calcul et la plus essentielle est de pouvoir travailler dans un environnement



réverbérant. C'est pour cette raison nous avons choisi cette méthode pour le reste de nos expériences [12][13][14].



**Figure 2.3 :** Schéma fonctionnel du GCC.

### 2.3.3 Formation de voies

La formation de voies (Beamforming) l'une des techniques utilisées pour les traitements des signaux à l'aide d'un réseau de capteur et elle est peut-être considérée aussi comme un filtre spatial pour chaque direction d'intérêt (l'antenne). Ce filtre va permettre de récupérer ou d'émettre des signaux dans des directions particulières avec un gain maximum, c'est-à-dire de favoriser le gain pour les directions d'intérêt et d'atténuer les signaux dans les directions non souhaitées (interférences). A la sortie de ce filtre, on n'aura que les signaux dans la direction d'intérêt. Domaine d'utilisation : les antennes intelligentes.

Dans une salle de conférence typique, le signal désiré provient de la source et est corrompu par le signal de bruit parasite avant d'atteindre les microphones. En exploitant la technique de formation de voies, le réseau de microphones tente d'obtenir un signal acoustique de haute qualité, en particulier dans les communications en champ lointain.

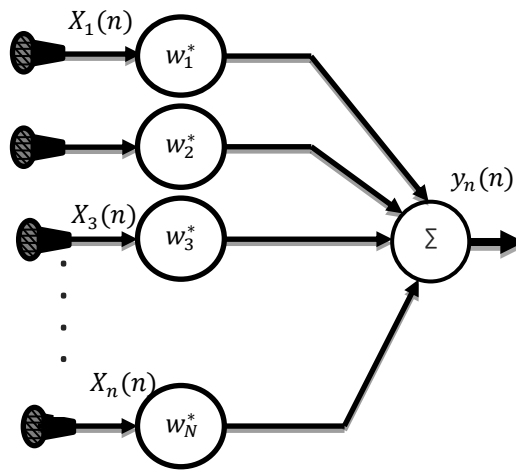
La formation de voies est utilisée dans une grande variété d'algorithmes de traitement de réseau qui nécessitent une capacité de capture du signal dans une direction particulière. Elle est également utilisée dans des applications de communication telles que les radars, les sonars et l'ingénierie médicale. La mise en forme de faisceau consiste à combiner la sortie d'un microphone pour obtenir un "faisceau" dans la direction qui nous intéresse. Ce faisceau fait de l'ensemble un microphone très directif. Les capteurs placés comme un réseau de microphones pour échantillonner spatialement une onde sonore ciblée sur eux. Toutes ces techniques dépendent de la directivité du signal souhaité.

### **Types de formateur de voies**

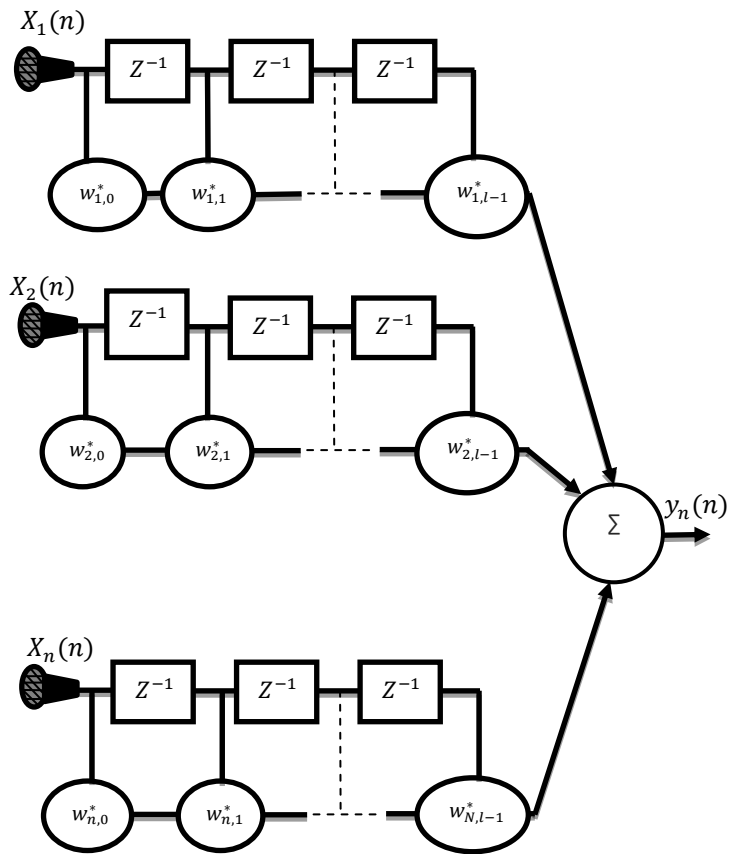
La formation de voies étant un outil très important dans le traitement des signaux de la localisation des sources sonores, il est essentiel de connaître les différents types ainsi que les techniques mises en œuvre et leurs avantages. Selon les types d'applications, on distingue deux types de réseau formateur de voies: le réseau d'analogique (Analog Beamforming) et le réseau numérique (Digital Beamforming).

Le premier type réalise l'algorithme sur un support Radiofréquence RF (ligne de transmission, guide d'onde...). D'autre côté, le deuxième type réalise la formation de voies sur un support numérique, beaucoup plus flexible. Ce type de formateur est la somme de deux technologies : la technologie d'antenne et la technologie numérique. Nous sommes intéressés dans ce mémoire pour expliquer la méthode de localisation par le deuxième type (**Structures des formateurs de voies numériques**).

Les algorithmes de la FV varient en fonction de la distance entre les sources et le réseau de récepteurs (microphones) par rapport à l'installation du réseau (le champ proche et le champ lointain). Dans les problèmes de traitement d'antenne classiques comme la gamme de navigation sonore (sonar), radar...etc. Les algorithmes de la FV ont aussi assumé que les sources sont localisées dans le champ lointain en vue de faciliter le traitement. La réponse fréquentielle d'un filtre RIF avec des poids  $w_i^*$  avec :  $1 \leq i \leq N$ .



(a)



(b)

**Figure 2.4.** : Structures des formateurs de voies : (a) modèle d'un champ proche (bande étroite), (b) modèle d'un champ lointain (bande large).

Une autre information qui doit être connue a priori pour la conception d'une formation de voies est la largeur de bande du signal source soit champ proche ou champ lointain car ceci a un grand impact sur le choix de sa structure. La figure (2.4) illustre deux formateurs de voies. Le premier, échantillonne l'onde dans l'espace est typiquement utilisé pour des signaux à bande étroite. La sortie  $y_n(n)$  à l'instant  $n$  est donnée par la combinaison linéaire des données reçues au niveau de chaque élément  $i$  à l'instant  $n$  :

$$y_n(n) = \sum_{i=1}^N w_N^* X_n(n) \quad (2.16)$$

On appelle un Champ lointain lorsque la distance entre la source sonore et le réseau de microphones est beaucoup plus grande que la taille du réseau, cette situation est appelée situation de champ lointain. Dans ce cas, les ondes sonores apparaissent pour être « planaire » en atteignant le système de la localisation. D'autre part lorsque la distance entre la source et le réseau est à peu près égale ou inférieure à sa taille, il s'agit de la condition de champ proche.

Si les sources sont assez près du réseau on dit qu'on est dans le champ proche, la courbure du front d'onde doit être prise en considération. Au contraire si les sources sont assez loin du réseau on dit alors qu'on est dans le champ lointain et les fronts d'onde au niveau du réseau peuvent être vus comme des ondes planes.

### L'entrée

La formation de voies avec la somme de retard pondérée unitaire qui a été brièvement présentée au chapitre 1 qui peut être créée en retardant les signaux de microphones  $x_m(t)$  avec le retard de direction appropriés  $\delta_m$  avec  $m=1,2,3...M$ . pour aligner dans le temps, puis en additionnant tous ces signaux alignés dans le temps, mathématiquement, il est défini comme suit :

$$y(t, \delta_1, \delta_2, \dots, \delta_M) \equiv \sum_{m=1}^{m=M} x_m(t - \delta_m) \quad (2.17)$$

Pour les signaux du microphone soient alignés dans le temps, les délais de pilotage  $\delta_m$  peuvent être réglés comme :

$$\delta_m = \tau_m - \tau_0 \quad (2.18)$$

où  $\tau_m$  est le délai entre la source et le microphone  $m$ ,  $\tau_0$  est réglé au minimum de tous les délais  $\tau_i$ ,  $i=[1,2,3,\dots,M]$  pour rendre  $\delta_m$  positif (système causal).

Nous pouvons maintenant exprimer la sortie de système par la somme de retards en termes de signal sonore de réponse impulsionnelle :

$$y(t, \delta_1, \delta_2, \dots, \delta_M) = s(t) * \sum_{m=1}^{m=M} h(\vec{d}_s, t - \tau_m + \tau_0) + \sum_{m=1}^{m=M} n_m(t - \tau_m + \tau_0) \quad (2.19)$$

Lorsque le filtre est appliqué à la formation de voies à la somme de retards, on obtient une formation de voies filtrée :

$$Y(W, \delta_1, \delta_2, \dots, \delta_M) = \sum_{m=1}^{m=M} G_m(\omega) X_m(\omega) e^{-j\omega\delta_m} \quad (2.20)$$

où  $X_m(\omega)$  est la transformée de fourier du signal de microphone  $x_m(t)$ , et  $G_m(\omega)$  est la Transformée de Fourier du filtre.

### **La sortie (SRP)**

La sortie d'un formateur de faisceau est connue sous le nom de réponse dirigée. Lorsque le point ou la direction de l'analyse correspond à l'emplacement source, le SRP sera maximisé. Toutefois les performances de localisation des techniques classiques de faisceaux de direction appliquent des filtres aux signaux de réseaux afin d'améliorer leurs performances. Quant à la phase, le filtre à transformer est incorporé avec la méthode du faisceau de direction, l'algorithme résultant (SRP-PHAT) est supérieur dans la lutte contre les effets néfastes du bruit de fond et de la réverbération par rapport à la méthode conventionnelle de faisceaux de direction et à la méthode par paire GCC-PHAT.

La figure (2.2) illustre la sortie de formateurs de voies. Le premier, échantillonne l'onde dans l'espace est typiquement utilisé pour des signaux à bande étroite. La

sortie  $R_{kl}$  à un instant  $\tau$  est donnée par la combinaison linéaire des données reçues au niveau de chaque élément:

En général, la puissance de réponse dirigée (SRP) est la puissance de sortie d'un beamformer à filtre et la somme de retards lorsqu'il dirige le système sur tous les points sur le même axe d'une région prédéfinie. Dans le domaine fréquentiel, elle est définie comme:

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \int_{-\infty}^{+\infty} Y(W, \delta_1, \delta_2, \dots, \delta_M) Y^*(W, \delta_1, \delta_2, \dots, \delta_M) d\omega \quad (2.21)$$

En remplaçant l'équation (2.20) par (2.21)

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \int_{-\infty}^{+\infty} \left( \sum_{A=1}^{A=M} G_A(\omega) X_A(\omega) e^{-j\omega\delta_A} \right) \left( \sum_{B=1}^{B=M} G_B^*(\omega) X_B^*(\omega) e^{-j\omega\delta_B} \right) d\omega \quad (2.22)$$

Donc,

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \int_{-\infty}^{+\infty} \sum_{A=1}^{A=M} \sum_{B=1}^{B=M} (G_A(\omega) G_B^*(\omega)) (X_A(\omega) X_B^*(\omega)) e^{-j\omega(\delta_A - \delta_B)} d\omega \quad (2.23)$$

A travers l'équation (2.18)

$$\delta_A - \delta_B = \tau_A - \tau_B \quad (2.24)$$

Et par l'insertion de (2.24) dans (2.23), nous obtenons :

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \int_{-\infty}^{+\infty} \sum_{A=1}^{A=M} \sum_{B=1}^{B=M} (G_A(\omega) G_A^*(\omega)) (X_B(\omega) X_B^*(\omega)) e^{-j\omega(\tau_A - \tau_B)} d\omega \quad (2.25)$$

Notez que l'intégrale converge car en pratique, les microphones et les filtres ont une énergie finie. Par conséquent, les sommations peuvent être échangées avec les intégrales comme suit :

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \sum_{A=1}^{A=M} \sum_{B=1}^{B=M} \int_{-\infty}^{+\infty} (G_A(\omega)G_B^*(\omega))(X_A(\omega)X_B^*(\omega))e^{-j\omega(\tau_A-\tau_B)}d\omega \quad (2.26)$$

Définir la fonction de pondération

$$\psi_{kl}(\omega) \equiv G_k(\omega) G_l^*(\omega) \quad (2.27)$$

Rappelons que l'équation x nous donne :

$$\tau_A - \tau_B = \tau_{AB} \quad (2.28)$$

En remplaçant les expressions (2.27) et (2.28) par (2.26), on obtient la fonction de SRP :

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \int_{-\infty}^{+\infty} \sum_{A=1}^{A=M} \sum_{B=1}^{B=M} (G_A(\omega)G_B^*(\omega))(X_A(\omega)X_B^*(\omega))e^{-j\omega\tau_{AB}}d\omega \quad (2.29)$$

Nous rappelons maintenant la fonction de GCC :

$$R_{AB}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi_{AB}(\omega)X_A(\omega)X_B^*(\omega) e^{j\omega\tau_{AB}}d\omega \quad (2.30)$$

On peut facilement constater que la SRP et le GCC ont des fonctions presque identiques, sauf que la SRP est additionnée sur tous les paires de microphones, et qu'il y a un décalage constant de  $2\pi$ . Cela donc nous donne un moyen de calculer la puissance de réponse dirigée SRP d'un réseau de microphones en additionnant la fonction de l'Intercorrélation GCC de toutes les paires dans un réseau de microphones (ici le décalage constant est ignoré puisqu'il ne s'agit d'un scalaire).

De façon similaire et dans le cas de la GCC-PHAT, lorsque la fonction de pondération (transformation de phase PHAT) est appliquée à la puissance de réponse dirigé (SRP), on obtient la SRP-PHAT pour chaque point de l'axe définie dans l'espace :

$$P(\delta_1, \delta_2, \dots, \delta_M) \equiv \sum_{A=1}^{A=M} \sum_{B=1}^{B=M} \int_{-\infty}^{+\infty} \frac{1}{|X_A(\omega)X_B^*(\omega)|} X_A(\omega)X_B^*(\omega) e^{j\omega\tau_{AB}} d\omega \quad (2.31)$$

Comme la GCC entre deux microphones, la même chose pour la méthode de SRP. Les éléments qui s'additionnent pour former la fonction de SRP-PHAT ci-dessus. Par conséquent, la partie de la SRP-PHAT change avec l'axe soit dans la partie supérieure ou inférieure de la matrice. En d'autre terme, pour le point particulier dans l'espace, la partie de la SRP-PHAT de l'équation (2.31) qui change peut-être calculée en additionnant la GCC de toutes les paires du réseau avec M-microphone.

Comme nous l'avons brièvement décrit dans la méthode de l'Intercorrélation, pour trouver le retard temporel de la source, nous dirigeons la formation de voies sur tous les points possible d'un volume vocal contenant la source sonore. Les points qui donnent de la sortie SRP-PHAT (the maximum weighted output power of SRP-PHAT) seront les emplacements de sources. Pour un seul signal de la source sonore, l'estimation de la direction de cette source sur un plan horizontal  $\vec{x}_s$  est :

$$\vec{x}_s = \text{Arg-max } P(\vec{x}) \quad (2.32)$$

où  $P(\vec{x})$  est la SRP-PHAT au point  $\vec{x}$  et est définie dans l'équation (SRP-PHAT).

L'hypothèse est que la fonction atteindra le maximum à l'emplacement réel de la source, même dans des conditions très bruyants et très réverbérant. Le problème avec SRP-PHAT est son cout de calcul élevé car l'espace de recherche comporte plusieurs réseaux de microphones au même temps [15][16][17][18][19].

### 2.3.4 Méthode à haute résolution (approches des sous-espaces)

La méthode haute résolution pour la détection et la localisation de sources par les techniques sous espaces s'appuient sur la décomposition spectrale de la matrice de covariance des signaux provenant des capteurs. Elles séparent l'espace caractéristique de la matrice de covariance du signal global en sous espace signal propre et sous espace bruit en utilisant une décomposition en valeurs propres de la matrice de covariance. La méthode MUSIC (Multiple Signal Classification) est un



exemple de ce type de méthodes qui a d'abord été proposée par Schimidt en 1986, pour estimer les directions d'arrivée des signaux. Cette méthode est capable de détecter plusieurs sources proches et présente de meilleures performances que la méthode de formation de faisceaux conventionnelle à condition que ces sources soient décorréélées ou faiblement corrélées. Si les sources sont fortement corrélées cette méthode nécessite une modification par les algorithmes de décorrélation [20].

Pour définir la méthode de la décomposition en valeurs singulières (SVD), il convient de commencer de SRP exprimé sous forme de matrice. Nous définissons le vecteur X

$$X=[X_{1,2}[0] \quad X_{1,2}[1] \quad \dots \quad X_{M-1,M}[N/2] ]^T \quad (2.33)$$

Avec  $X \in \mathbb{C}^{P \left( \frac{N}{(2+1) \times 1} \right)}$

De même, la matrice  $W \in \mathbb{C}^{Q \times P(N/(2+1))}$  contient tous les Coefficients SRP-PHAT

$$W = \begin{bmatrix} W_{1,1,2}[0] & W_{1,1,2}[1] \dots & W_{1,M-1,M} \left[ \frac{N}{2} \right] \\ \vdots & \ddots & \vdots \\ W_{Q,1,2}[0] & W_{Q,1,2}[1] \dots & W_{Q,M-1,M} \left[ \frac{N}{2} \right] \end{bmatrix} \quad (2.34)$$

Enfin, le vecteur  $Y \in \mathbb{C}^{Q \times 1}$  stocke l'énergie SRP-PHAT pour toutes les sources potentielles Q et est obtenu à partir de ce qui suit multiplication matricielle :

$$Y = [ Y \dots YQ ]^T = R\{WX\} \quad (2.35)$$

Comme mentionné pour SRP, cette multiplication matricielle est coûteuse en termes de calcul alors que les possibilités sont nombreuses les positions des sources à scanner. Pour faire face à cette limitation, nous proposons d'effectuer une décomposition de la valeur singulière sur la matrice W, où  $U \in \mathbb{C}^{Q \times K}$ ,  $S \in \mathbb{C}^{K \times K}$  et  $V \in \mathbb{C}^{P(N/(2+1)) \times K}$ , comme indiqué au point (2.19).

$$W \approx USV^H \quad (2.36)$$

Où  $\{\dots\}^H$  représente l'opérateur hermitien.

Le paramètre  $\in N \cap ]0, Kmax]$ , où la valeur supérieure  $Kmax = \max \{Q, P \left( \frac{N}{2+1} \right)\}$ , est choisie pour garantir une reconstruction précise de  $W$ , selon l'état dans (2.20), où le paramètre  $\delta$  défini par l'utilisateur est un petit positif qui modélise l'erreur de reconstruction tolérable. L'opérateur  $Tr \{...\}$  représente la trace de la matrice.

$$Tr\{SS^T\} \geq (1 - \delta)Tr\{WW^H\} \quad (2.37)$$

Le vecteur  $Z \in \mathbb{C}^{K \times 1}$  résulte de la projection des observations  $X$  dans le sous-espace des dimensions  $K$  :

$$Z = V^H X \quad (2.38)$$

La matrice  $D \in \mathbb{C}^{Q \times K}$  est obtenue en (2.22) et peut être décomposée en un ensemble de vecteurs  $Q$   $Dq \in \mathbb{C}^{1 \times K}$ :

$$D = US = [D_1^T \ D_2^T \ \dots \ D_Q^T]^T \quad (2.39)$$

L'indice de la DOA la plus probable obtenu maintenant correspond à :

$$\bar{q} = \operatorname{argmax}\{R\{Dq \cdot Z^H\}\} \quad (2.40)$$

Une façon pour trouver la valeur correcte de  $q$  dans (2.40) en calculant chaque  $Yq$  pour  $q \in N \cap [1, Q]$ , puis en trouvant l'indice  $q$  qui conduit à la valeur maximale, ce qui évidemment implique un nombre important de calculs, car la complexité L'ordre est linéaire ( $O(Q)$ ). Il est donc pertinent de regarder pour une autre fonction de coût qui permettrait une recherche. Pour toutes les valeurs de  $q$ , les expressions  $\|Dq\|_2$  sont presque identiques (lorsque la reconstruction remplit la condition (2.37) pour une petite valeur de  $\delta$ ), mais n'est pas nécessairement égale à 1. Nous avons donc défini les nouveaux vecteurs  $\hat{D}q = Dq / \|Dq\|_2$  et la valeur normalisée vecteur  $\hat{Z} = Z / \|Z\|_2$ . Avec  $\|\hat{Z}\|_2^2 = 1$ , et  $\|\hat{D}q\|_2^2 = 1$ , Le produit des points peut donc être exprimé comme suit :

$$R\{Dq \cdot Z^H\} + 1 - \frac{1}{2} \|\hat{D}q - \hat{Z}^H\|_2^2 \quad (2.41)$$

Et donc maximiser (2.40) correspond maintenant à minimiser  $\|\hat{D}q - \hat{Z}^H\|_2^2$ . Cette

minimisation peut être réalisée par le calcul  $\|\widehat{D}q - \widehat{Z}^H\|_2^2$  pour toutes les valeurs de  $q$  et de trouver  $q$  qui mène à la valeur minimale, mais cela ramène à la complexité linéaire ordonner  $O(Q)$  comme dans (2.40). Heureusement, la nouvelle formulation basée sur la somme des carrés devient une recherche du plus proche voisin qui peut être résolu efficacement en utilisant un arbre k-d. L'algorithme 1 résume la configuration hors ligne effectuée avant le traitement et les calculs en ligne. Les performances en temps réel sont indépendantes des performances de calcul l'UDS et la construction d'arbres, qui sont coûteux, car ils sont hors ligne. Lors du traitement en ligne, le calcul du vecteur  $Z$  implique un ordre de complexité de  $O(KPN)$  et l'arbre k-d. La recherche présente en moyenne une complexité  $O(\log Q)$  [21].

## 2.4 Matériels : Microsoft Kinect Xbox 360

### 2.4.1 Définition

Initialement connu sous le nom code Project natal, est un périphérique destiné au matériel Microsoft avec la Xbox 360 pour le v1 et la Xbox 1 et Windows depuis le v2 qui ont été lancés en septembre 2008 permettant de contrôler une interface sans utiliser sa manette dans les jeux vidéo.

La Kinect xbox360 (voir la figure) de Microsoft a été lancée en novembre 2010 comme un accessoire. Elle a été développée par la société PrimeSense en collaboration avec Microsoft. Le site était destiné à offrir une toute nouvelle façon d'interagir avec la console par des gestes et la voix à la place du contrôleur traditionnel.



Figure 2.5 : La Kinect XBOX 360.

## 2.4.2 Caractéristique

Capteur :

- Lentilles détectant vocale et la profondeur.
- Micro a reconnaissance vocale.
- Capteur motorisé pour suivre les déplacements des objets.

Champs de visions

- Champ de vision horizontal : 57 degrés.
- Champ de vision vertical: 43 degrés.
- Marge de déplacement du capteur :  $\pm 27$  degrés.
- Portée du capteur : 1.2 mètre jusqu'à 3.5 mètre.

Flux de données :

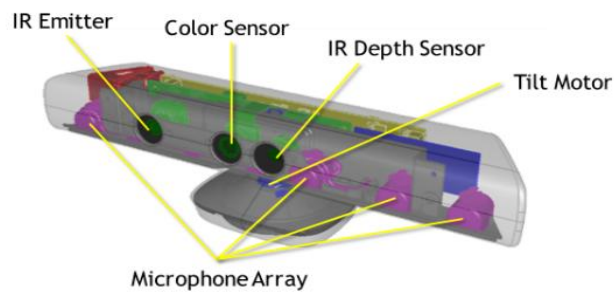
- 320\*240 en couleur 16 bits à 30 images par seconde.
- 640\*480 en couleur 32 bits à 30 images par seconde.
- Audio 16 bits à 16 kHz.

Système de reconnaissance physique :

- Jusqu'à 6 personnes et 2 joueurs actifs.
- 20 articulations par squelette.
- Application des mouvements des joueurs sur leurs avatars Xbox live.

Audio :

- Chat vocal Xbox live et chat vocal dans les jeux vidéo.
- Suppression de l'écho.
- Reconnaissance vocal multilingue.



**Figure 2.6 :** illustration des composants de la Kinect.

Le schéma fonctionnel du modèle de référence est présenté à la figure (2.6). Le Primesense Le système sur puce (SoC) PS1080 fournit une image de profondeur synchronisée, une image couleur et les flux audio. Tous les algorithmes d'acquisition de la profondeur fonctionnent sur le SoC PS1080, donc aucune charge de calcul n'est ajoutée à l'hôte. Les fonctions de niveau supérieur comme la scène L'analyse et le suivi doivent être effectués sur l'hôte. Un dispositif de détection similaire utilisant la conception de référence Primesense est disponible auprès d'Asus.

Caméra RGB : La caméra RGB du Kinect dispose d'une résolution standard de 640x480 pixels et fonctionne à une fréquence d'image de 30 Hz. Un mode haute résolution offrant 1280x1024 pixels peut être utilisé. Mais la fréquence d'images tombe à environ 15 Hz en mode haute résolution. La sortie native de la caméra RGB est codée en tant que motif Bayer2 image, mais les peuvent convertir l'information brute en une image RVB standard.

Capteur de profondeur : Le capteur de profondeur est constitué d'un laser infrarouge de longueur d'onde 830 nm qui projette une de points sur son champ de vision. Une caméra infrarouge enregistre ces motifs sur les objets et un processeur de signal numérique (DSP) embarqué calcule la distance en corrélant l'image en direct avec les motifs de référence stockés.

La sortie du capteur de profondeur est un flux vidéo de 640x480 pixels. Chaque pixel contient 11 bits d'informations de profondeur, donc 2048 (2<sup>11</sup>) différentes valeurs pour représenter la distance sont possibles. Pour des raisons de calcul

interne, les huit colonnes de pixels les plus à droite dans l'image ne contiennent pas de données. La résolution utilisable est donc réduite à 632x480. Les bibliothèques permettant d'accéder au flux de profondeur Kinect prennent en charge la conversion des valeurs brutes de profondeur 11 bits aux coordonnées du monde réel en millimètres. En raison du déplacement horizontal du projecteur infrarouge et de la caméra infrarouge, le projecteur et la caméra ne partagent pas exactement le même champ de vision. Par conséquent, un l'ombre, en particulier pour les objets proches de la caméra, est visible.

Réseau de microphones : Le capteur Kinect comprend un réseau de microphones à 4 canaux avec ADCs3 fonctionnant à 16 kHz taux d'échantillonnage et résolution de 24 bits. En raison de sa fréquence d'échantillonnage plutôt faible et donc de sa fréquence de coupure d'environ 8kHz, il est principalement destiné à l'enregistrement de la parole. Les algorithmes peuvent être utilisés pour séparer plusieurs haut-parleurs devant le capteur Kinect. Cependant, ces algorithmes ne sont pas fournis par l'appareil lui-même. Les quatre microphones à électret omnidirectionnels sont situés derrière une grille sur le fond de la barre de capteurs. La Kinect permet de recevoir un flux audio à quatre canaux pour effectuer l'annulation de l'écho ce qui permet d'obtenir un meilleur rapport signal/bruit [22] [23].



**Figure 2.7** : Réseau de microphones dans une Kinect Xbox 360.

### 2.4.3 Domaine d'application de Kinect

Depuis la sortie de Kinect par Microsoft, la précision et la stabilité des données de Kinect telles que la camera de profondeur et les microphones, ont été des éléments essentiels et importants dans la recherche et l'analyse des données. Afin de développer des moyens efficaces pour analyser le dispositif lui-même et l'utilisation de plusieurs données. Les chercheurs ont besoin de données de profondeur (depth data) de haute qualité pendant l'étape de prétraitement, qui est très décisif pour la précision des résultats. L'une des principales préoccupations des chercheurs est d'éliminer le bruit des images et de convertir ces images pour obtenir des meilleures qualités. D'autre coté, les microphones de la Microsoft Kinect sont largement utilisés pour le traitement de signal parole (speech enchantement, localisation d'une source sonore, réduction de bruit...etc.).

#### Applications :

- Principalement et largement utilisé à des fins de jeu.
- "Cabine d'essayage " virtuelle.
- la vidéo conférence.
- la navigation sur le web.



#### **L'utilisation dans les hôpitaux**

- Kinect présente également un potentiel d'utilisation convaincant en médecine
- Kinect pour l'examen préopératoire, de l'imagerie médicale, permettant au chirurgien d'accéder aux informations sans contamination.



### Vers l'avenir

- L'armée américaine fait de son mieux pour utiliser cette technologie dans l'invention des drones, c'est-à-dire l'Ariel sans pilote et d'autres véhicules de l'armée.
- Fins médicales.
- Des robots de vigilance artificielle, des vaisseaux spatiaux pour l'analyse d'autres plantes et bien plus encore....



**Figure 2.8 :** Véhicule aérien sans pilote (UAV).



## **2.5 Conclusion**

Les stratégies de la localisation des sources acoustiques étaient bien décrites dans ce chapitre tout en citant les différentes classes et la classification adoptée repose sur le nombre de microphones disponibles. Dans le prochain chapitre nous allons présenter la configuration et la partie expérimentale de la localisation d'une source sonore qui correspond à notre environnement.

# Chapitre 3 Localisation en 2D par l'algorithme

## GCC -PHAT

---

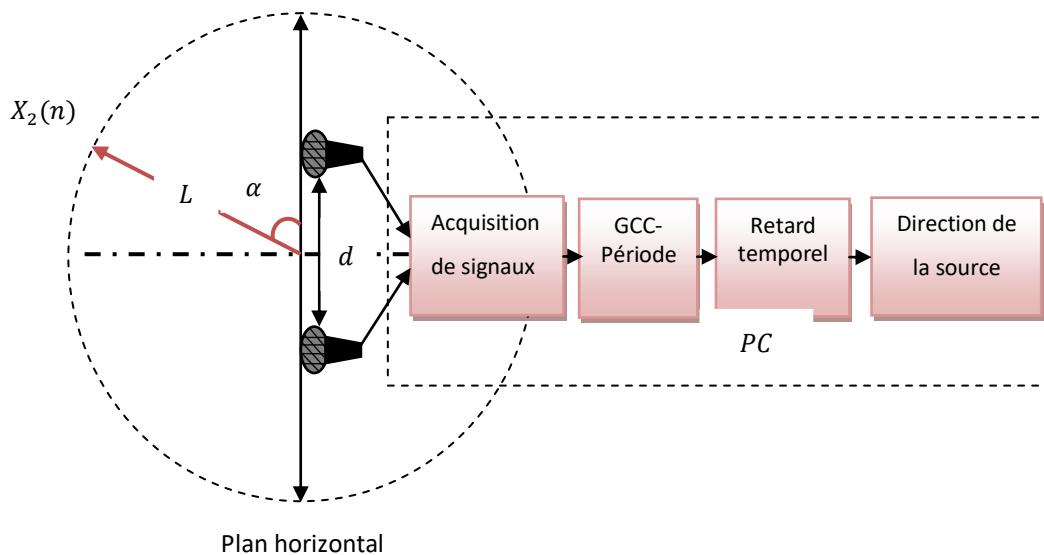
### 3.1 Introduction

La localisation d'une source sonore active peut être réalisée par des réseaux de microphones dans différents environnements (réverbérations). L'objectif de ce mémoire est de localiser une source en deux dimensions en se basant sur un réseau de quatre microphones par l'algorithme GCC-PHAT. Pour cela, deux microphones sont installés dans un plan horizontal séparés par une distance donnée. La source sonore a été générée par un téléphone portable. En utilisant les quatre microphones du Kinect, on peut déterminer les coordonnées de la source en 2D. Au cours de ce chapitre, la conception de ce système sera réalisée et testée sous l'environnement MATLAB qui nous permettra l'acquisition de données et leurs prétraitements via la Kinect Xbox du Microsoft.

### 3.2 Description du système

Le schéma de la figure (3.1) décrit d'une manière globale les principales fonctions du système de localisation de direction d'arrivée de la source qu'on désire réaliser dans la première partie. Ce système est constitué de deux microphones pour capter les signaux audio et un ordinateur pour l'acquisition des signaux et l'exécution des algorithmes.

Les expériences sont réalisées dans un espace dégagé est préservé. Les distances entre microphones seront justifiées au cours de ce chapitre.



**Figure 3.1:** Description globale du système.

### 3.4 Localisation de la source (la direction d'arrivée)

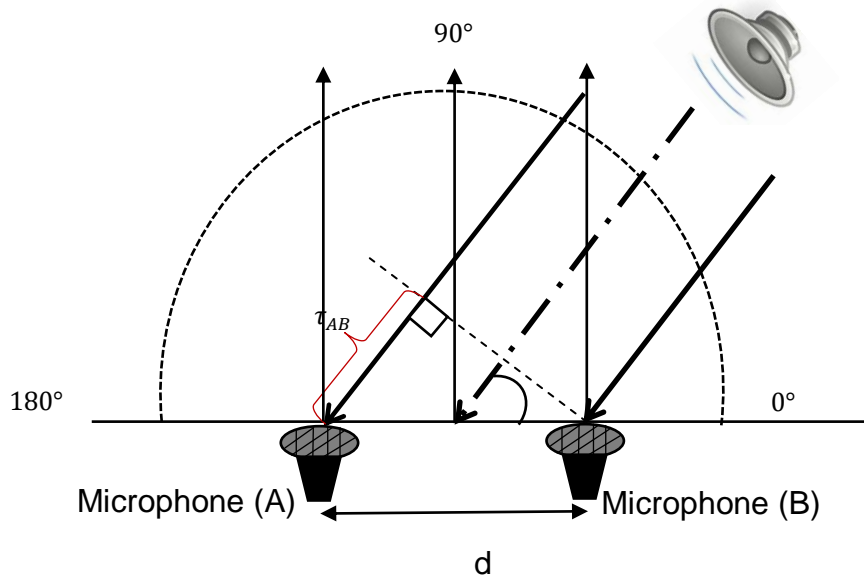
Dans un environnement acoustique, la localisation de sources joue un rôle très important dans des applications diverses par exemple : la poursuite automatique du locuteur par une caméra dans une salle de vidéoconférence. La localisation comprend trois dimensions spatiales, soit l'azimut (gauche droite), le plan vertical (haut-bas), et la distance. Ici, uniquement le plan azimutal qui est pris en considération.

D'après les méthodes de localisation citées dans le premier chapitre, les méthodes d'estimation du retard temporel et les principes d'estimation semblent les plus simples et possibles à employer.

#### 3.4.1 Principe

Soit, selon la figure (3.2), deux microphones omnidirectionnels soumis à une onde acoustique plane progressive sous un angle d'incidence tel qu'un front d'onde atteint d'abord le premier, puis le second après un temps de propagation  $\tau$  ; le second

microphone capte la même forme d'onde que le premier, mais avec un retard. Par un raisonnement géométrique, on constate que  $\tau$  a une relation avec l'angle d'incidence  $\alpha$  et la distance entre les microphones.



**Figure 3.2:** principe de l'estimation de l'angle d'arrivée d'une source sonore.

Les ondes sonores apparaissent pour être planaire en atteignant le système de la localisation.

A travers deux microphones on peut estimer la source par la relation:

$$\alpha = \cos^{-1} \left( \frac{\tau_{AB} * C}{d} \right) \quad (3.1)$$

Avec  $C$  est la célérité du son,  $\tau$  est le retard temporel. La mesure du retard temporel permet de déterminer l'angle d'incidence. Si la distance entre source sonore et le réseau de microphone est beaucoup plus grande que la taille du réseau on l'appelle un champ lointain.

### 3.4.2 La résolution angulaire

La résolution angulaire peut être définie comme étant la capacité d'un système de discriminer la source sonore et la précision de l'estimation. Elle est déterminée par le nombre de différentes mesures de directions d'arrivées entre 0 et  $\pi$ .

Considérons le même réseau de microphones précédent. Pour une distance entre les microphones, on a le temps de parcours de cette distance  $\tau$  :

$$\tau = \frac{d}{c} \quad (3.2)$$

Ce temps de parcours  $\tau$  équivaut un nombre d'échantillons  $N$ , tel que :

$$N = \tau * F_s \quad (3.3)$$

Le nombre  $N$  correspond aussi au nombre des angles pouvant être détectés.

$$N = \frac{d}{c} * F_s \quad (3.4)$$

D'après l'équation (3.4), l'amélioration de la résolution angulaire peut être achevée par les deux méthodes suivantes:

- Augmenter la distance entre les microphones  $d$  ; Ceci implique une taille importante du réseau, ce qui n'est pas toujours adéquat surtout dans des situations où l'espace est limité. Aussi, une grande distance peut causer des problèmes qui a une grande influence dans le contexte et la performance de la méthode.
- Augmenter la fréquence d'échantillonnage ; Ceci augmente le nombre d'échantillons constituant le retard  $\tau$ , d'où chaque échantillon correspond à un angle d'arrivée, mais ceci influence aussi sur la qualité d'estimation et la rapidité du système.

Dans la figure (3.3), la variation de la résolution angulaire en fonction de quatre fréquences d'échantillonnage est illustrée. Pour une distance  $d= 0.4$  m,  $F_s=8$ khz. On obtient 9 échantillons et pour un 16 kHz, 18 échantillons.

### 3.5 La localisation en deux dimensions

Dans cette section, on s'intéresse à la localisation d'une source en 2D. L'algorithme GCC -PHAT (Generalized Cross Correlation PHASE Transform) est utilisé et testé dans un environnement réel. Cet algorithme est basé sur le calcul de l'intercorrélacion entre les deux signaux reçus. La fonction de pondération PHAT (un filtre de blanchiment) est employée pour rendre le pic indiquant le retard plus étroit donc facile à détecter.

Dans le cas de la localisation en deux dimensions, lorsque deux spatialement récepteurs séparés sur le même plan (axe des x) détectent un signal, la différence du temps mesuré d'arrivée sert à localiser la direction de la source, et lorsque deux paires de récepteurs sont utilisées (deux microphones sur l'axe des x et deux microphones sur l'axe des y), la source peut être localisée par l'intersection de deux hyperboles.

#### 3.5.1 Définition de l'hyperbole

Une hyperbole est le lieu géométrique de tous les points dont la différence des distances à deux points fixes appelés foyers est constante. Cette constante est égale à la distance entre les deux sommets de l'hyperbole. Le point milieu du segment joignant les deux sommets est le centre de l'hyperbole [24].

$$|d(P, F1) - d(P, F2)| = \text{Constante.}$$

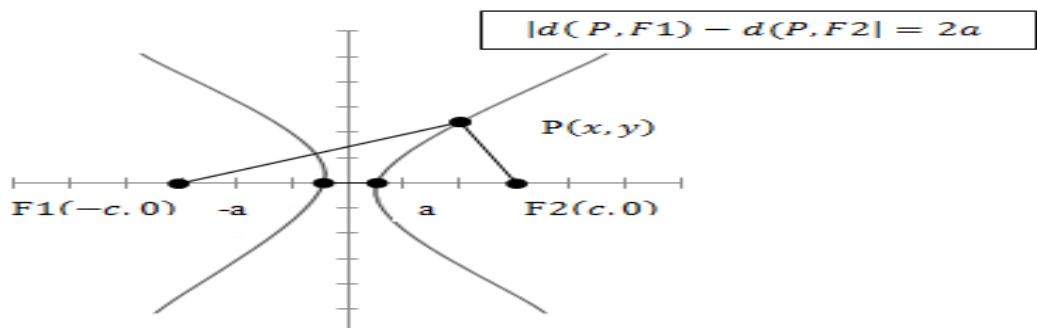


Figure3.3 : une fonction hyperbolique.

### 3.5.2 Caractéristiques de l'hyperbole

- L'hyperbole possède deux sommets.
- L'hyperbole possède deux foyers.
- L'hyperbole possède deux asymptotes.
- La droite passant par les deux foyers est l'axe transversal et la droite passant par le point d'intersection entre les asymptotes et perpendiculaire à l'axe transversal est l'axe conjugué.
- On peut former un rectangle délimité par les sommets de l'hyperbole et les asymptotes.

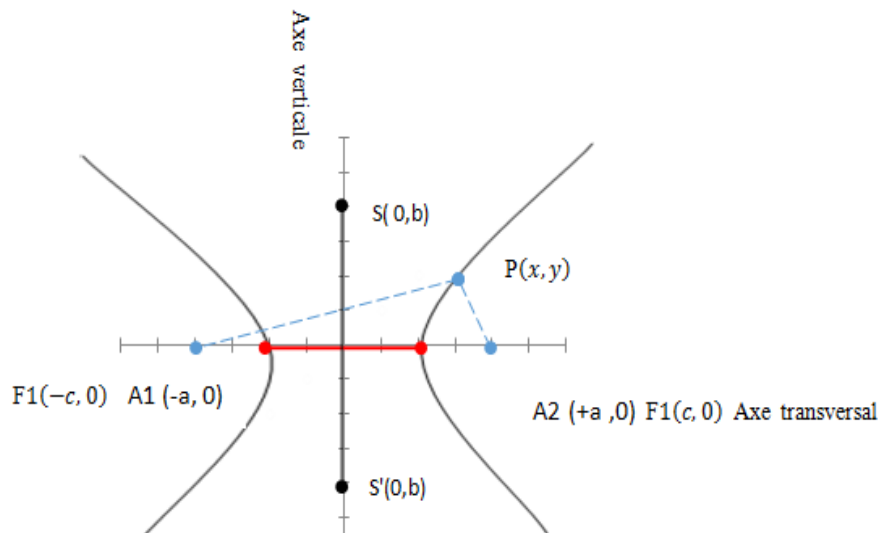


Figure 3.4 : l'hyperbole centrée à l'origine.

### 3.5.3 L'équation de l'hyperbole

L'hyperbole H est l'ensemble des points  $P(x ; y)$  de  $\mathbb{R}^2$

$$H = \left\{ (x; y) \mid \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \right\} \subset \mathbb{R}^2 \quad (3.5)$$

#### a. L'hyperbole horizontale

$$\frac{(x-h)^2}{ax^2} - \frac{(y-k)^2}{bx^2} = 1 \quad (3.6)$$

$a$  est la mesure entre un sommet de l'hyperbole et son centre,  $b$  est la moitié de la

hauteur du rectangle de l'hyperbole, (h, k) sont les coordonnées du centre de l'hyperbole (intersection des asymptotes).

b. L'hyperbole verticale

$$\frac{(y-k)^2}{by^2} - \frac{(x-h)^2}{ay^2} = 1 \quad (3.7)$$

Avec « a » est la moitié de la largeur du rectangle de l'hyperbole, « b » est la mesure entre un sommet de l'hyperbole et son centre. (h, k) sont les coordonnées du centre de l'hyperbole (intersection des asymptotes).

**3.5.4 Relations dans les fonctions hyperboliques**

**Tableau 3.1:** Relation entre hyperbole de base et transformée.

	Hyperbole verticale		Hyperbole horizontale	
Définition	Tous les point dont la différence des distances à deux points fixes (foyers) est constante=2b		Tous les points dont la différence des distances à deux points fixes (foyers) est constante = 2a	
Asymptotes	Pente = $\pm \frac{b}{a}$			
Hyperbole de base	Sommets	S : (0,b) S' : (0,-b)	Sommets	A1 : (a, 0) A2 (-a, 0)
	Foyers	$c^2 = a^2 + b^2$ F1 : (0,c), F2 (0,-c)	Foyers	$c^2 = a^2 + b^2$ F1 : (c,0), F2 (c,0)
	Rectangle	R1 : (a,b) R2 : (a,-b)	R3: (-a,-b) R4:( -a,b)	/
Hyperbole transformée	Sommets	S: (h,k+b) S': (h,k-b)	Sommets	S: ( h+a,k) S': ( h-a,k)
	Foyers	F1: (h,k+c) F2: (h,k-c)	Foyers	F1: (h+c,k) F2: (h-c,k)
	Rectangle	R1 : (h+a ,k+b) R2 : (h+a,k-b)	R3: (h-a,k-b) R4:( h-a,k+b)	/



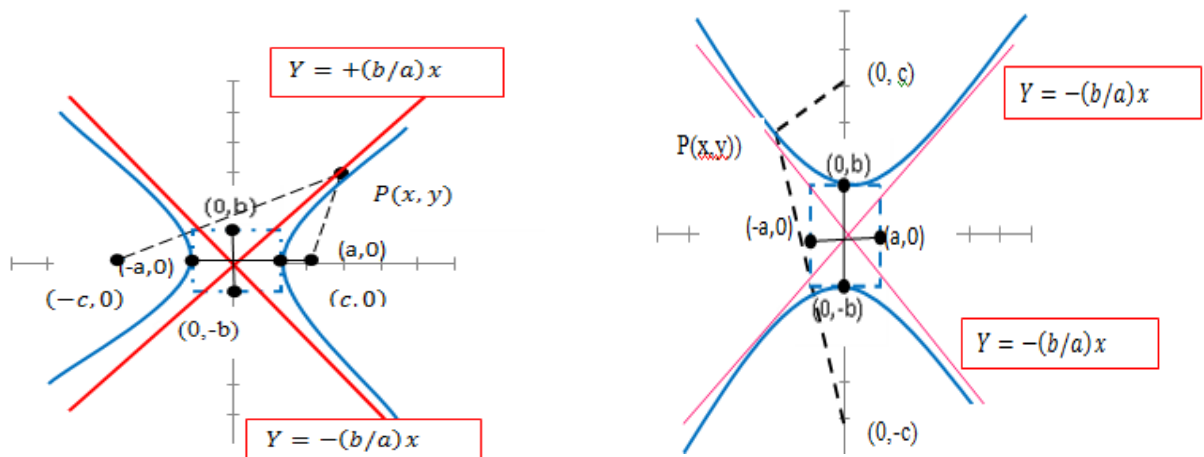


Figure 3.5 : l'hyperbole a l'axe horizontal et verticale.

### 3.5.5 L'équation d'une hyperbole à partir d'un graphique

Afin de déterminer d'une hyperbole à partir d'un graphique, il faut déterminer la valeur des différents paramètres  $a$ ,  $b$  et  $h$ ,  $k$ . On peut faire une liste de stratégies qui s'avèrent souvent utiles :

- Déterminer la valeur des paramètres  $h$  et  $k$  à partir des coordonnées du centre de l'hyperbole.
- Si l'hyperbole est horizontale, déterminer la valeur des paramètres  $a$  qui correspond à la distance entre le centre et l'un des sommets de l'hyperbole.
- Si l'hyperbole est verticale, déterminer la valeur du paramètre  $b$  qui correspond à la distance entre le centre et l'un des sommets de l'hyperbole.
- Déterminer la valeur de  $a, b$  ou  $c$  à l'aide des coordonnées du foyer, du rectangle de l'hyperbole, d'un point sur l'hyperbole et/ou de la relation Pythagore.
- Choisir la bonne forme d'équation selon l'orientation de l'hyperbole et y remplacer la valeur des paramètres.
- Déterminer l'équation des asymptotes.

### 3.5.6 Tracer une hyperbole à l'aide de l'équation

Afin de tracer une hyperbole à l'aide de l'équation. On peut suivre les étapes suivantes :

- Identifier les paramètres h et k dans l'équation.
- Tracer asymptotes.
- Déterminer les sommets de l'hyperbole, puis tracer l'esquisse.

### 3.5.7 Partie expérimentale

Dans la figure (3.6) quatre microphones sont placés dans un réseau rectangulaire, le microphone mic (1) et le microphone mic (2) sont installés sur l'axe des X en  $(cx, 0)$  et  $(-cx, 0)$ , le microphone mic (3) et microphone mic (4) sont installés sur l'axe Y en  $(0, cy)$  et  $(0, -cy)$  respectivement. Ces microphones sont séparés par une distance  $R$  avec  $R = 2M_i$  [25] .

La distance entre la source  $(x_s, y_s)$  et chaque microphones est donné par :

$$\begin{cases} d_1 = \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2} \\ d_2 = \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2} \\ d_3 = \sqrt{(x_3 - x_s)^2 + (y_3 - y_s)^2} \\ d_4 = \sqrt{(x_4 - x_s)^2 + (y_4 - y_s)^2} \end{cases} \quad (3.8)$$

Les  $(X_i, Y_i)$  ( $i=1, \dots, 4$ ) sont les coordonnées des microphones et  $(x_s, y_s)$  sont les coordonnées du son source par rapport à l'origine au centre du réseau. En supposant  $M_1 = R/2$  est la distance entre l'origine et le mic (1) ou mic (2) , et  $M_2$  est la distance entre l'origine et le mic (3) ou mic(4). En conséquence, nous pouvons récrire (3.8) comme suit :

$$\begin{cases} d_1 = \sqrt{(M_1 - x_s)^2 + y_s^2} \\ d_2 = \sqrt{(-M_1 - x_s)^2 + y_s^2} \\ d_3 = \sqrt{x_s^2 + (M_2 - y_s)^2} \\ d_4 = \sqrt{x_s^2 + (-M_2 - y_s)^2} \end{cases} \quad (3.9)$$

A partir de x-microphones, le mic (1) sera le premier à enregistrer le son provenant de la source et le mic (2) capte le même signal audio avec une temporisation. La différence de temps de chaque paire de microphones est donnée

par : 
$$\tau_i = \frac{d_i}{v} \quad (3.10)$$

Donc 
$$\begin{cases} d_x = \tau_x * v \\ \text{et} \\ d_y = \tau_y * v \end{cases} \quad (3.11)$$

Alors : 
$$\begin{cases} \tau_x = \frac{d_x}{v} \\ \text{and} \\ \tau_y = \frac{d_y}{v} \end{cases} \quad (3.12)$$

Où  $dx = d2 - d1$  et  $dy = d4 - d3$ .

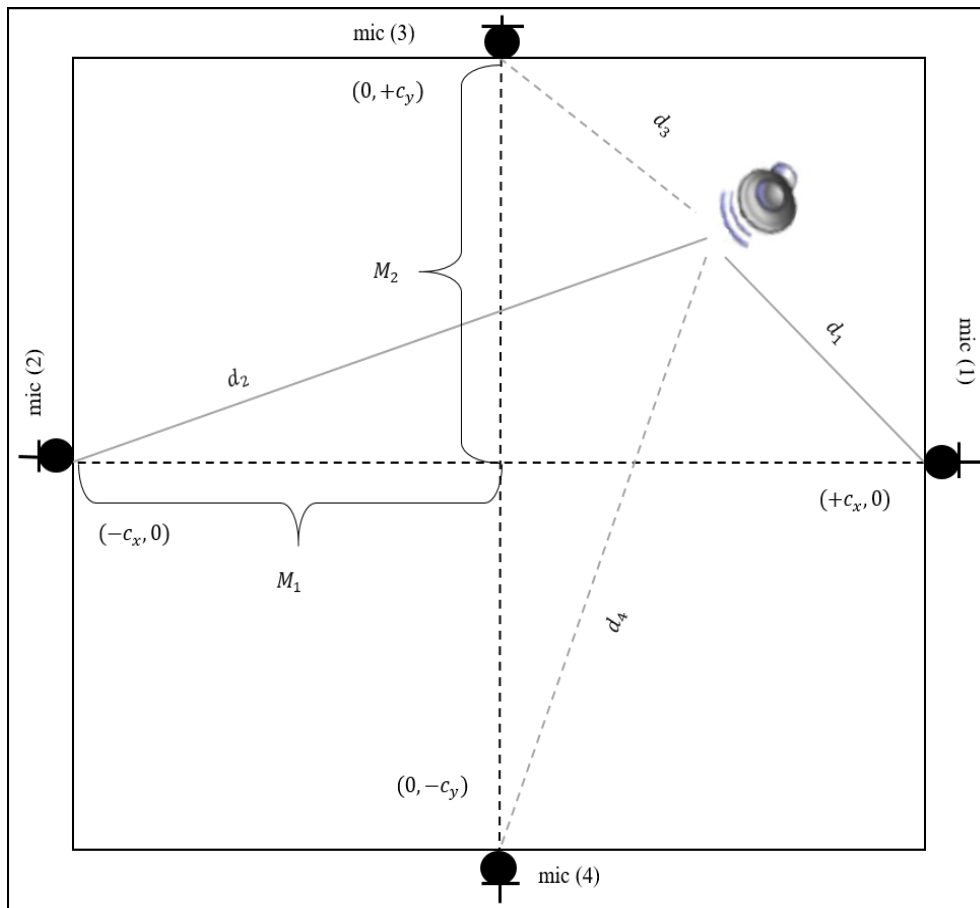


Figure 3.6 : Illustration de la configuration de la simulation.

A partir des coordonnées du microphone X, l'équation pour une hyperbole dans une norme est donnée par ce qui suit équation :

$$\frac{x^2}{a_x^2} - \frac{y^2}{b_x^2} = 1 \quad (3.13)$$

Cette équation décrit une hyperbole centrée à l'origine avec une ouverture gauche-droite avec des foyers à  $(\pm c_x, 0)$ . La distance du mic(1) au mic(2) est  $2a_x$  et il est égal à  $\tau_x * v$ . En utilisant les sommets  $(\pm a_x, 0)$  de l'hyperbole, on peut calculer les différences de la distance entre le point  $(x_s, y_s)$  et les deux x-microphones par  $a_x = \tau_x * v/2$  et en utilisant (3.9). D'après l'équation des foyers  $c_x^2 = \sqrt{a_x^2 + b_x^2}$  nous pouvons obtenir les extrémités de l'axe conjugué comme  $b_x = \sqrt{c_x^2 - a_x^2}$ . En outre, hyperbole centrée à l'origine (0,0) avec une ouverture haut-bas est donnée comme suit :

$$\frac{y^2}{a_y^2} - \frac{x^2}{b_y^2} = 1 \quad (3.14)$$

En remplaçant (3.13) dans (3.14), la position du son peut être localisée par les coordonnées  $(x_s, y_s)$ :

$$\begin{cases} x_s = a_x b_y \sqrt{\frac{a_y^2 + b_x^2}{b_x^2 b_y^2 + a_x^2 a_y^2}} \\ y_s = a_y b_x \sqrt{\frac{a_x^2 + b_y^2}{b_x^2 b_y^2 + a_x^2 a_y^2}} \end{cases} \quad (3.15)$$

### **3.6 Conclusion**

Dans cette partie, l'algorithme de localisation PHAT-GCC basé sur l'estimation du retard temporel TDOA est choisi parce qu'il est simple en termes d'implémentation et temps de calcul, convenable pour les implémentations en temps réel et applicable dans les différents environnements, pour le but d'estimer la position d'une source sonore de signal acoustique en utilisant une console Microsoft Kinect et des microphones comme récepteurs. Les quatre microphones sont séparés les uns des autres et sont connectés à la console Kinect qui servira d'interface entre les captures et le programme MATLAB. L'expérience consistera à l'aide d'un papier millimètres à mesurer la position de la source sonore.

## Chapitre 4 Résultat et Discussions

---

### 4.1 Introduction

Dans ce chapitre nous allons implémenter la méthode de localisation de source basée sur la différence de temps d'arrivée (TDOA) et nous allons la tester dans trois expériences en utilisant deux microphones dans les deux premières expériences et un réseau de quatre microphones de la Microsoft Kinect.

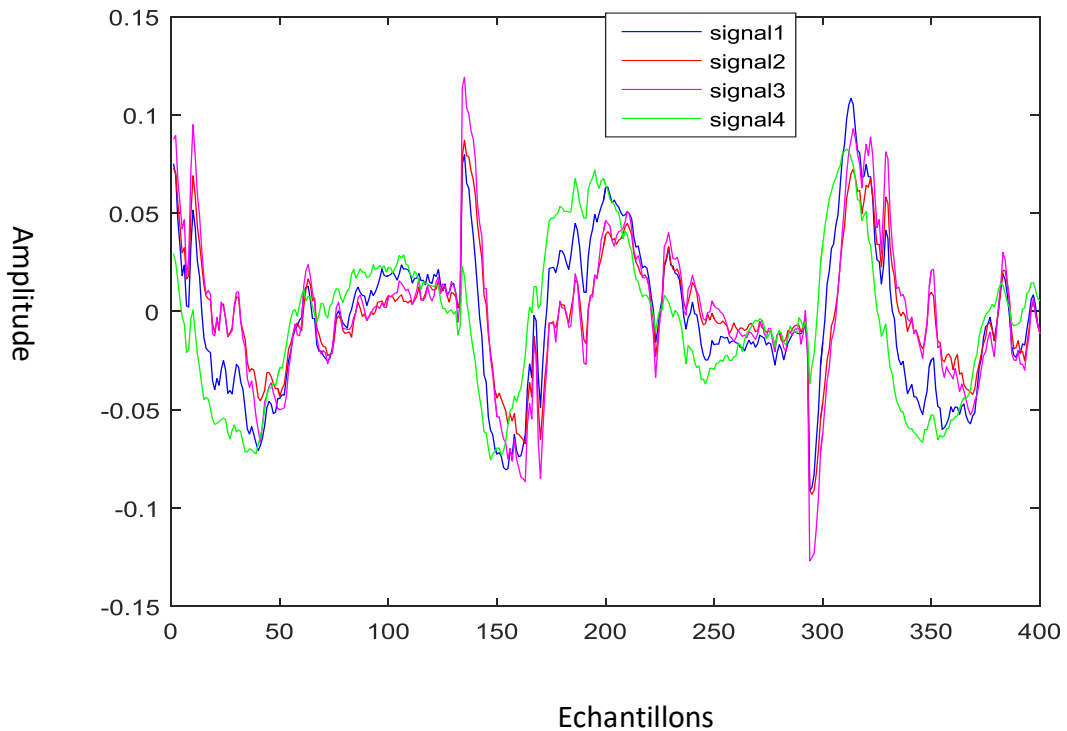
### 4.2 Expérience 1

Les caractéristiques des différents éléments utilisés sont représentées dans les tableaux (4.1) ci-dessous :

Élément	Caractéristique
La source sonore : Haut-parleur	téléphone portable
Microphone	Microsoft Kinect Xbox 360
L'environnement (une salle)	4 x4 x 2.8 m <sup>3</sup>
Objects	Armoire, 2 Lits, table, chaise, bureau

**Tableau 4.1:**Caractéristiques et informations des éléments utilisées.

La figure (4.1) présente les signaux acquis par les 4 microphones.

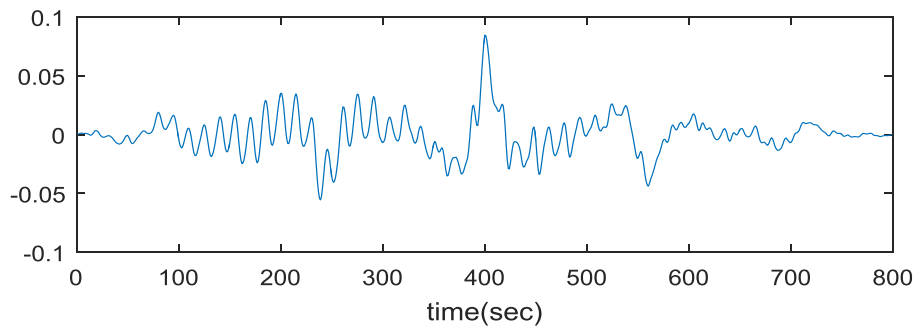


**Figure4.1** : Les signaux reçus par les microphones.

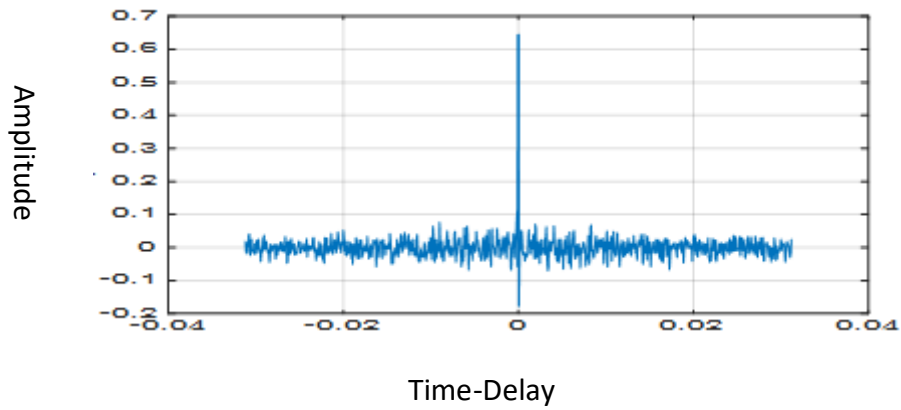
Dans cette expérience 2 microphones et un haut-parleur sont disposés comme dans la figure (4.4). La fréquence d'échantillonnage est de 16 kHz et la résolution 8 bits. L'expérience consistera à mesurer la distance  $D$  et pour chaque valeur de  $D$  varier la position du haut-parleur en variant la valeur de  $\alpha$  entre  $(0^0$  et  $180^0)$  avec une distance de la source =  $D$ .

Ensuite pour chaque valeur de  $\alpha$ , nous utilisons l'algorithme de détection basé sur TDOA développé ici afin de d'estimer la valeur de  $\alpha$  et donc la position du haut-parleur et enfin de calculer l'erreur absolue entre la valeur  $\alpha$  mesurée de l'angle et sa valeur estimée  $\alpha_e$  cette erreur est donnée par la relation  $\Delta\alpha = |\alpha - \alpha_e|$ .

Comme spécifié à la relation (3.1) La détermination de  $\alpha$  passe par le calcul de la différence de temps entre les paires de microphones qui est obtenu en déterminant le point d'abscisse maximal de la courbe d'intercorrélation donc un exemple est donné à la figure (4.2) .et exemple pour l'intercorrélation généralisée à la figure (4.3) .



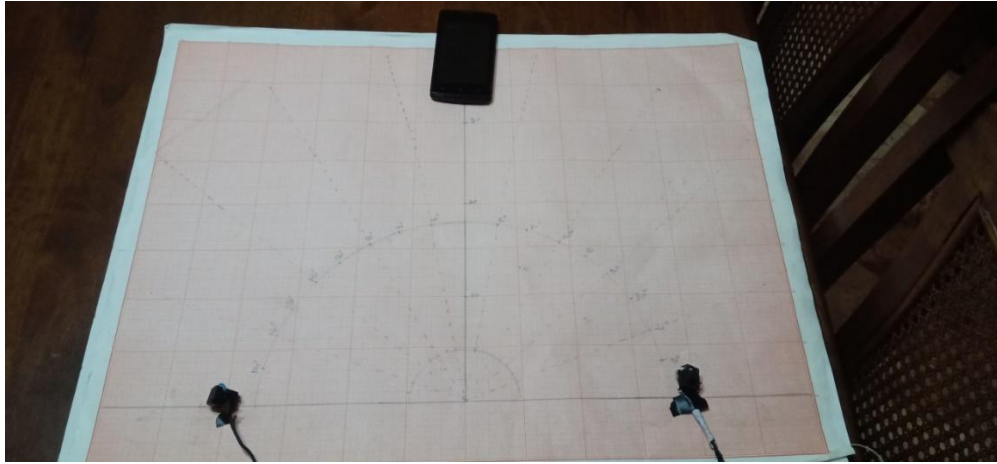
**Figure4.2** : plot d'intercorrélation (xcorr) en temps réel.



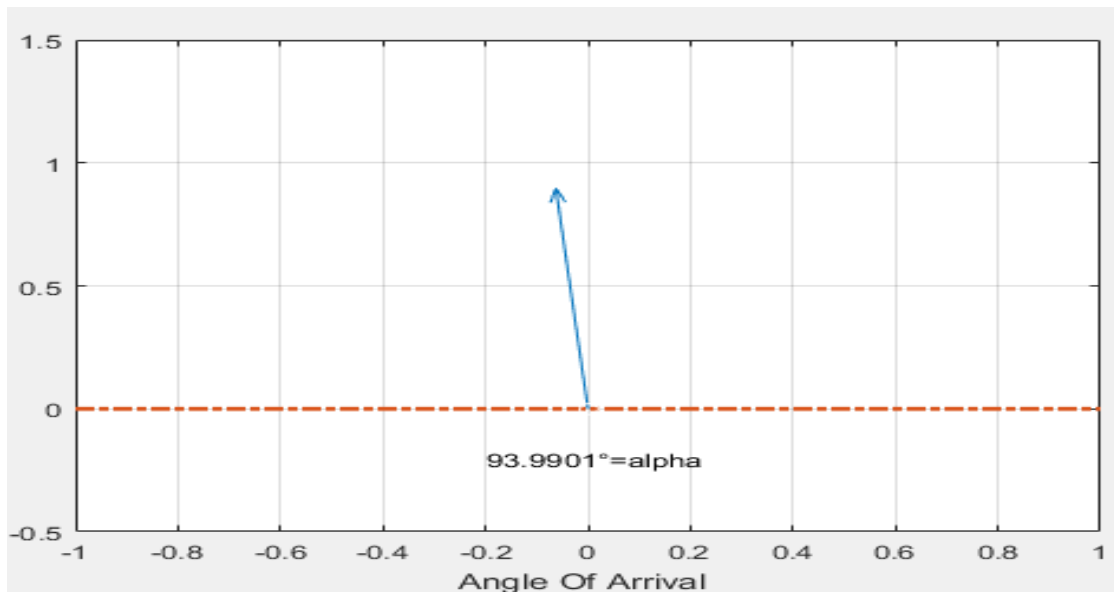
**Figure4.3** : l'intercorrélation généralisée (GCC\_PHAT).



Figure (4.4) présente une photo réelle de notre expérience et figure (4.5) le résultat sur matlab :



**Figure4.4** : Exemple d'expérimentation de l'expérience 1.



**Figure4.5** : L'estimation de la direction d'arrivée sous Matlab.

Résultat :

<b>d(m)</b>	<b>Angle Donné (dégrée)</b>	<b>Angle Estimé (dégrée)</b>	<b>Erreur (dégrée)</b>
<b>0.3</b>	0°	3,05	<b>3,05</b>
	20°	23,23	<b>3,23</b>
	40°	45,90	<b>5,9</b>
	60°	65,32	<b>5,32</b>
	80°	86,003	<b>6,003</b>
	90°	93,96	<b>3,96</b>
	100°	102	<b>2</b>
	120°	119,35	<b>0,65</b>
	140°	139,94	<b>0,06</b>
	160°	154,76	<b>5,24</b>
	180°	172,94	<b>7,06</b>
<b>0.5</b>	0°	3.62	<b>3,62</b>
	20°	25,29	<b>5,29</b>
	40°	54,23	<b>14,23</b>
	60°	67,58	<b>7,58</b>
	80°	82,91	<b>2,91</b>
	90°	94,78	<b>4,78</b>

	100°	106,79	<b>6,79</b>
	120°	115,62	<b>5,48</b>
	140°	131,97	<b>8,03</b>
	160°	156,70	<b>4,30</b>
	180°	180	<b>0</b>
<b>1</b>	0°	16,02	<b>16,02</b>
	20°	31,38	<b>11,38</b>
	40°	52,39	<b>12,39</b>
	60°	66,40	<b>6,40</b>
	80°	79,19	<b>1,81</b>
	90°	91	<b>1</b>
	100°	102,60	<b>2,60</b>
	120°	117,19	<b>3,81</b>
	140°	132,59	<b>8,41</b>
	160°	153,84	<b>7,16</b>
	180°	163,79	<b>16,21</b>

**Tableau 4.1:** L'estimation de l'angle d'arrivée avec la variation de la distance entre les microphones.

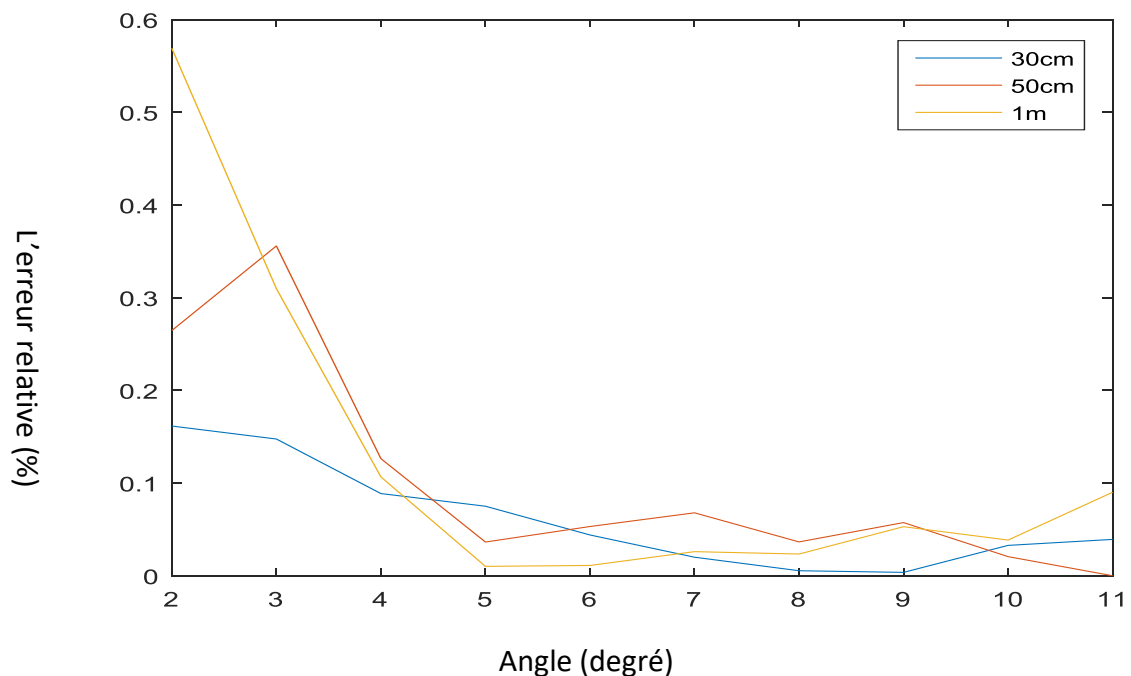
Observation :

Le tableau (4.2) représente les résultats de l'expérience. On constate que pour :

- D = 30 cm l'erreur est comprise entre 0.06 et 7.06 et qu'elle atteint sa valeur la plus faible pour  $\alpha = 140^{\circ}$
- D = 50 cm l'erreur est comprise entre 0 et 14.23 et qu'elle atteint sa valeur la plus faible pour  $\alpha = 180^{\circ}$
- D = 100 cm l'erreur est comprise entre 1 et 16.21 et qu'elle atteint sa valeur la plus faible pour  $\alpha = 90^{\circ}$

La relation qu'on peut tirer de ces observation entre d et la performance est que la performance de localisation décroît lorsque la valeur de D augmente, le système est donc plus précis a de courte distance.

La figure (4.6) présente une comparaison des erreurs entre les valeurs estimes et mesures obtenu à différente valeur de D pour différentes valeurs de  $\alpha$ .

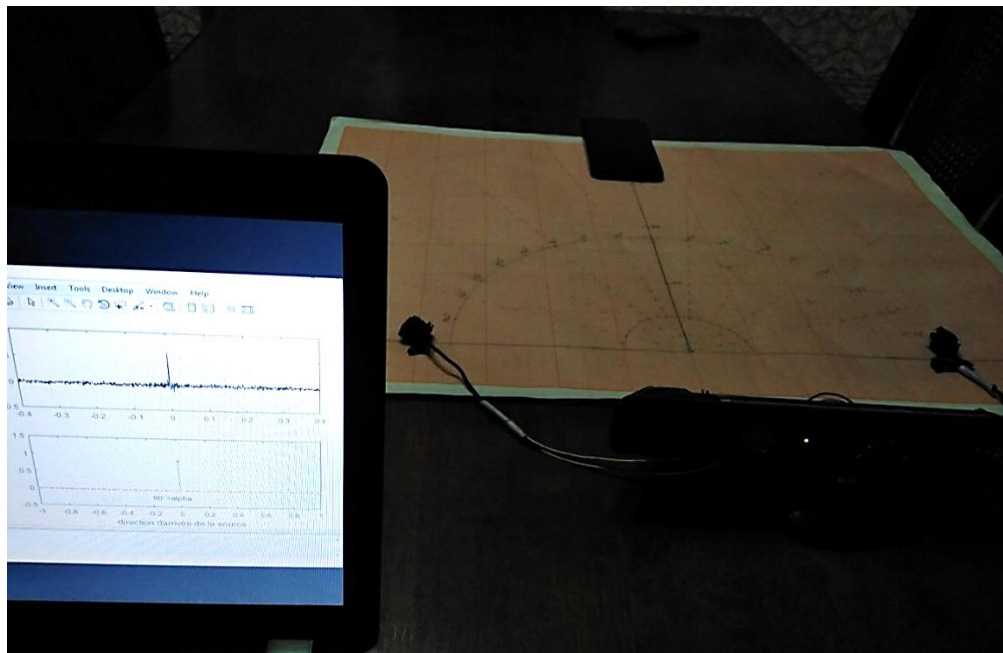


**Figure4.6 :** Comparaison des erreurs entre les valeurs D.

Nous avons observé que La meilleure distance entre les microphones pour localiser la source est de 0.3 m pour une distance de la source sonore de 0.3 m.

### 4.3 Expérience 2 :

L'expérience 2 consiste à placer le haut-parleur a la distance  $d=0,4$  m et faire varier la fréquence d'échantillonnage ( $F_s$ ) des signaux acquis via la console Kinect par notre programme Matlab afin d'estimer la direction de la source (haut-parleur) en utilisant la méthode présentée dans le chapitre 4. Figure (4.7) présente une photo réelle de notre expérience.



**Figure4.7** : Exemple d'expérimentation de l'expérience 2.

Résultat :

Fs (KHz)	Angle Donné (dégrée)	Angle Estimé (dégrée)	Erreur (dégrée)	Fs (KHz)	Angle Donné (dégrée)	Angle Estimé (dégrée)	Erreur (dégrée)
<b>1</b>	0°	33,18	<b>33,18</b>	<b>12</b>	0°	13,05	<b>13,05</b>
	20°	38,33	<b>18,33</b>		20°	25,29	<b>5,29</b>
	40°	63,88	<b>23,88</b>		40°	45,34	<b>5,34</b>
	60°	89,54	<b>29,54</b>		60°	73	<b>13</b>
	80°	104,1	<b>24,16</b>		80°	86,72	<b>6,72</b>
	90°	127,5	<b>37,52</b>		90°	93,33	<b>3,33</b>
	100°	146,6	<b>46,61</b>		100°	109,9	<b>9,92</b>
	120°	146,6	<b>26,61</b>		120°	112,9	<b>7,05</b>
	140°	178,7	<b>38,75</b>		140°	139,9	<b>0,01</b>
	160°	180	<b>20</b>		160°	154,7	<b>5,24</b>
180°	180	<b>0</b>	180°	169,9	<b>10,06</b>		
	0°	20	<b>20</b>		0°	7,44	<b>0</b>
	20°	41,22	<b>21,22</b>		20°	24,98	<b>3,02</b>
	40°	53,95	<b>13,95</b>		40°	47,27	<b>2,73</b>
	60°	73,02	<b>13,02</b>		60°	61,95	<b>2,55</b>
	80°	94,64	<b>14,64</b>		80°	84,99	<b>2,21</b>

<b>4</b>	90°	102	<b>12,04</b>	<b>16</b>	90°	92,53	<b>2,36</b>
	100°	102,8	<b>2,84</b>		100°	102,0	<b>2,88</b>
	120°	109,0	<b>10,96</b>		120°	118,1	<b>5,67</b>
	140°	129,6	<b>10,37</b>		140°	136,7	<b>3,95</b>
	160°	146,6	<b>13,39</b>		160°	152,5	<b>3,40</b>
	180°	180	<b>0</b>		180°	172,9	<b>0</b>
<b>8</b>	0°	16,02	<b>16,02</b>	<b>20</b>	0°	0	<b>0</b>
	20°	36,52	<b>16,52</b>		20°	16,98	<b>3,02</b>
	40°	47,08	<b>7,08</b>		40°	37,27	<b>2,73</b>
	60°	77,33	<b>17,33</b>		60°	57,45	<b>2,55</b>
	80°	95,99	<b>15,99</b>		80°	82,21	<b>2,21</b>
	90°	102,7	<b>12,71</b>		90°	87,64	<b>2,36</b>
	100°	114,8	<b>14,89</b>		100°	97,12	<b>2,88</b>
	120°	128,2	<b>8,24</b>		120°	125,6	<b>5,67</b>
	140°	136,9	<b>3,07</b>		140°	136,0	<b>3,95</b>
	160°	159,9	<b>0,6</b>		160°	156,6	<b>3,40</b>
	180°	159,9	<b>20,06</b>		180°	180	<b>0</b>

**Tableau 4.3:** L'estimation de l'angle d'arrivée avec la variation de  $F_s$ .

### Observation

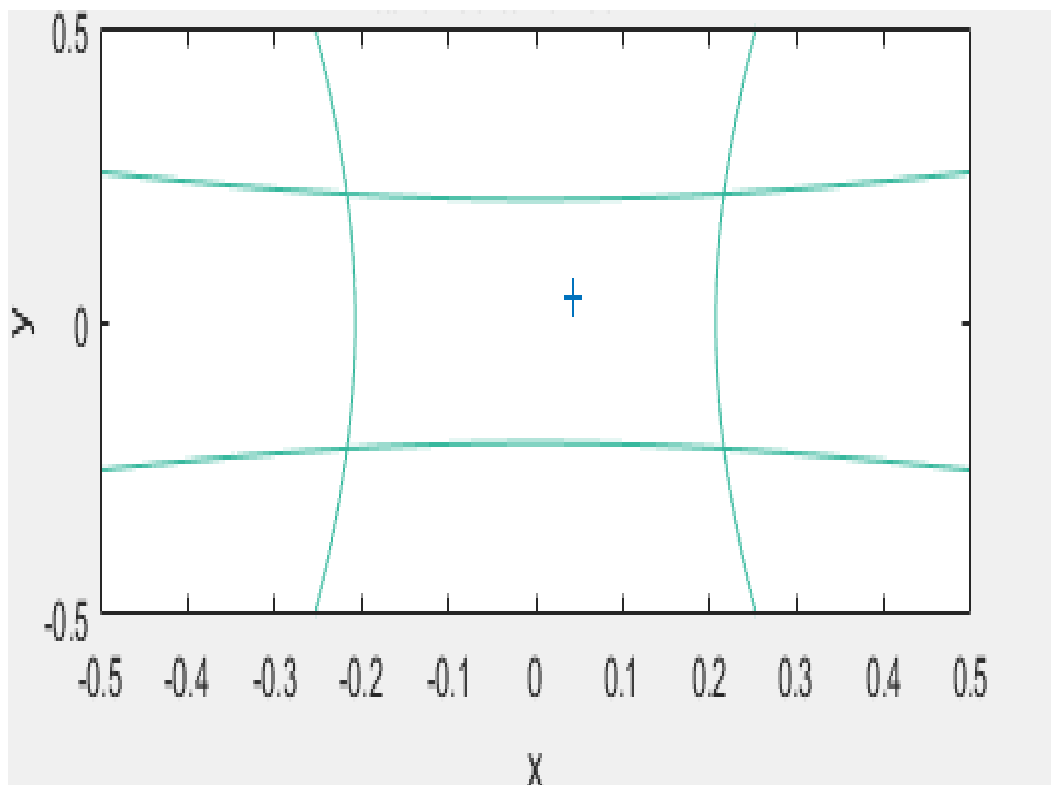
Augmente le nombre d'échantillons par la fréquence d'échantillonnage comme nous avons montré par l'équation (3.4) dans le chapitre 3 constituant le retard  $\tau$ , d'où chaque échantillon correspond à un angle d'arrivée, mais ceci influence négativement sur la rapidité du système.

On constate que le système est le plus précis lorsque la fréquence d'échantillonnage à la plus grande valeur à savoir 20kHz et l'erreur est plus grande lorsque sa valeur est réglée sur une valeur faible (1kHz). On peut ainsi conclure d'après le tableau (4.2) que plus la fréquence d'échantillonnage n'est pas élevée, plus le système n'est pas précis. En attribuant le temps d'arrivée, il a été possible de déterminer le temps avec une résolution d'environ une période d'échantillonnage (1/8000 s) pour  $F_s=8\text{kHz}$  et (1/16000 s) pour  $F_s=16\text{kHz}$ . Cela indique que les erreurs de la localisation dues au temps d'arrivée mesuré. Une estimation approximative de l'erreur de position attendant est donnée par  $\tau * v$ , Avec une fréquence d'échantillonnage de 8 kHz l'erreur de synchronisation est  $1/F_s=0.000125$  seconde ce qui conduit à une estimation de  $\tau * v = 0.000125 * 343 = 4.3$  cm.



#### 4.4 Expérience 3 :

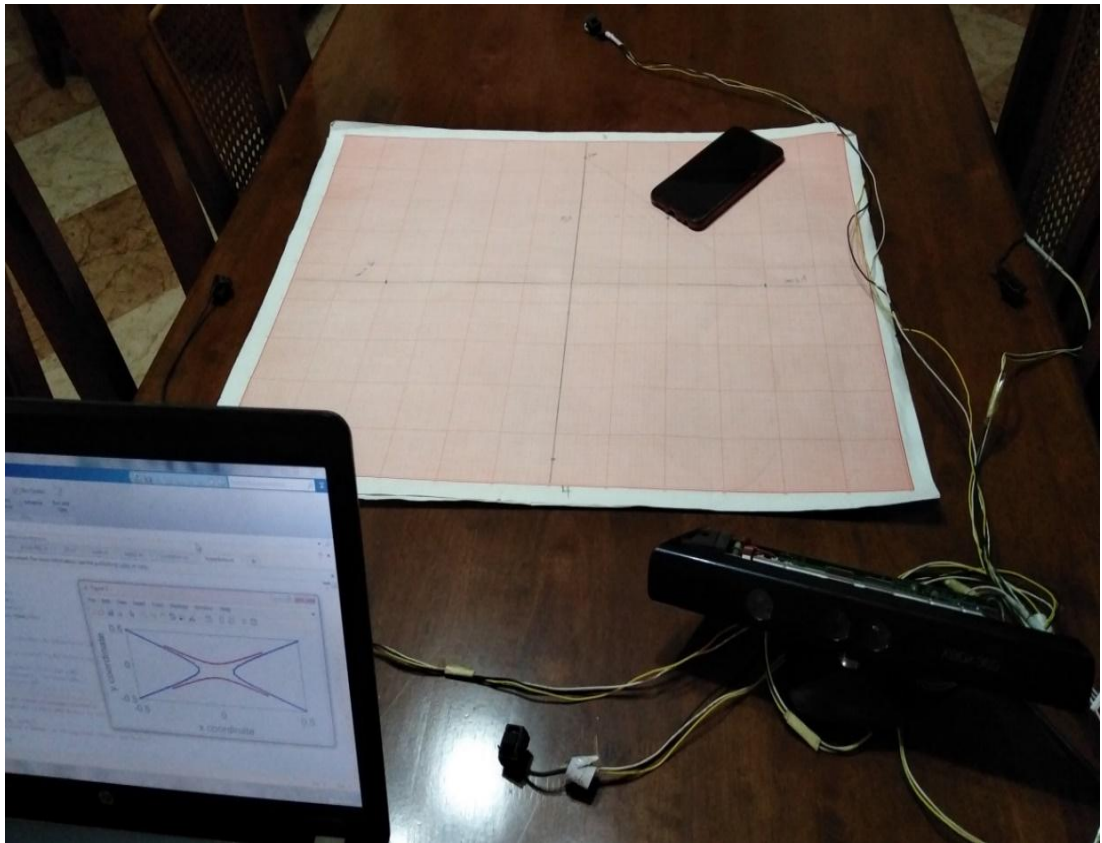
L'expérience 3 consiste à estimer la position d'une source sonore en deux dimensions. Le dispositif expérimental est le même que celui de l'expérience 1 à la différence qu'ici 2 microphones supplémentaires sont ajoutés, les 4 microphones sont disposés comme dans la figure (4.9). Lorsque le son est détecté, la TDOA est calculée et les hyperboles données par les relations (3.13) et (3.14) sont calculées et tracées (figure 4.8) :



**Figure4.8 :** La tendance des deux hyperboles.

La position estimée de la source est alors calculée en utilisant les paramètres  $(a_x, y)$  et  $(b_x, y)$  de l'intersection des 2 hyperboles. Les coordonnées de la source sont données par la relation (3.15).

L'expérience a été effectuée pour 9 positions du haut-parleur différentes dont les coordonnées sont présentes sur le tableau (4.4). Les résultats de l'expérience sont présentés à la figure (4.10). Avec toutes ces sources "réelles", le début n'était pas clair, et la forme précise de l'onde sonore arrive de plusieurs courses et manière. De bien, les meilleurs résultats ont été obtenus en enregistrant une courte impulsion sonore et en jouant avec un téléphone portable. Cette méthode fonctionne très bien parce que : (i) La source Sonore est très claire (ii) la source est reproductible (iii) l'hautparleur du téléphone est petit ce qui permet bien de définir la source.



**Figure4.9:** Exemple d'expérimentation de l'expérience 3.

Résultats

X (proposé)	0,1	0,1	0,3	0,4	0,4	0,3	0,4	0,4	0
Y (proposé)	0,1	0,3	0,1	0,1	0,2	0,3	0,3	0,4	0,5
X (estimé)	0,1	0,1	0,26	0,4	0,38	0,27	0,38	0,37	0
Y (estimé)	0,1	0,28	0,13	0,12	0,2	0,3	0,28	0,36	0,49

Tableau 4.4: Les coordonnées proposée pour la source sonore.

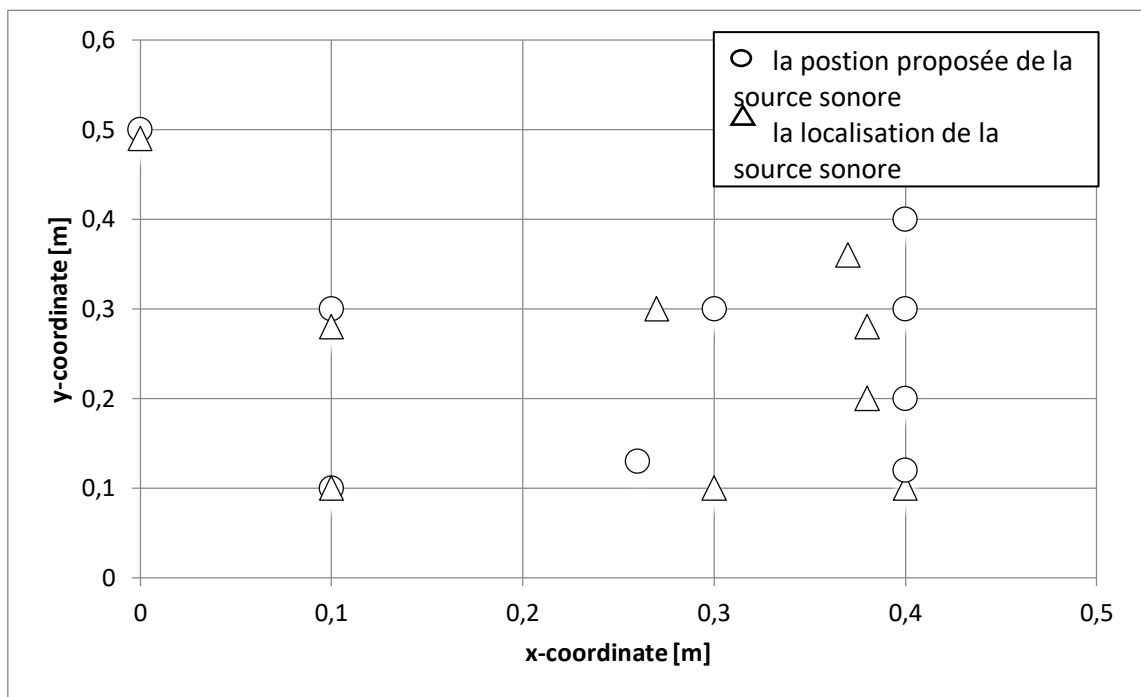


Figure 4.10 : la localisation de la source sonore en deux dimensions.

### Observation

Les résultats montrent qu'au point (0,0.5) on a une erreur minimale sur x nul et une erreur de 1.1 cm sur y, au point (0.3, 0.3) on a une erreur maximale sur x de 2.3 cm et nul sur y et enfin au point (0.4, 0.3) on a une erreur maximale sur y de 2 cm et une sur x de 2 cm. On remarque ainsi que la précision de la position de la source estimée par notre méthode augmente lorsque cette dernière est proche de l'origine et qu'elle se dégrade au fur et à mesure qu'on s'en éloigne.

## **4.5 Conclusion**

Dans ce chapitre une approche basée sur la différence de temps d'arrives a été appliquées afin de localiser une source de son dans un environnement donné. L'approche a été testée dans une première 2 expérience en variant certains paramètres tels que la distance de la source sonore et la fréquence d'échantillonnage. Ensuite elle a été testée dans 3<sup>ème</sup> expérience afin de localiser une source sonore dans un environnement réel. Les résultats obtenus ont été plutôt concluants et ont permis de démontrer l'efficacité de la méthode pour localiser avec précision une source sonore cette précision est d'autant plus grande lorsque l'on se rapproche de l'origine.

## Conclusion générale

---

L'objectif dans ce travail était de réaliser un système de localisation d'une source sonore basé sur la méthode de l'intercorrélation généralisée avec transformation de phase GCC-PHAT. L'idée principale de cette technique consiste à calculer la différence de temps d'arrivée (TDOA) entre les paires de microphones afin de déterminer soit la direction de la source (angle d'arrivée) ou bien l'endroit de la source par plusieurs relations mathématiques.

Pour réaliser cela, la connaissance des directions d'arrivées des sources est primordiale d'où la nécessité d'un algorithme de localisation des sources acoustiques. Pour tester la performance du système en temps réel, le dispositif de la Microsoft Kinect XBOX 360 était utilisé comme une carte d'acquisition responsable au prétraitement des signaux et leurs restitutions.

La réalisation de ce système en temps réel et dans un environnement acoustique réel a rencontré des limites que ce soit dans l'acquisition de données. L'acquisition et le traitement des signaux réels se fait généralement par un matériel (soft et hard). Cependant, lors de ce mémoire, l'idée a été réalisée par un matériel faible coût disponible dans tous les Établissements basé uniquement sur un micro-ordinateur, quatre microphones et un hautparleur. La carte d'acquisition utilisée n'est rien d'autre que la carte son fournie avec tous les micro-ordinateurs; elle semble être un moyen très intéressant pour le test des algorithmes dédiés au traitement de signal (dans la gamme audio). La carte son intégrée permet d'utiliser plus de deux microphones, ce qui permet d'utiliser plusieurs algorithmes que ce soit dans la localisation ou dans la formation de voies qui nécessitent plus de deux microphones pour leurs fonctionnements.

L'algorithme de localisation utilisé est basé sur l'estimation de la différence en retard temporel (TDOA). Ça implique deux étapes; dans la première étape le TDOA entre un pair de microphones est estimé, et dans la deuxième étape la position de la source est estimée à partir de ce TDOA et la géométrie du réseau.

Cette procédure a l'avantage d'être applicable dans tous les signaux acoustiques. La résolution de la localisation peut être ajustée d'une manière flexible en faisant varier la fréquence d'échantillonnage et ou la dimension du réseau de microphones.

Les méthodes d'intercorrélation dite généralisés GCC sont très efficaces en terme de temps de calcul (grâce à l'emploi de la FFT) par rapport à l'intercorrélation classique en plus qu'ils n'ont pas besoin de faire une recherche sur toutes les angles possibles. Elles prennent un temps de décision très court, donc elles ont la capacité d'une bonne poursuite de la source.

# Annexes

---

## Les microphones

### 1 Introduction

Un des principaux défis en robotique est de développer à l'instar de l'homme des capteurs qui recevraient sans discrimination les informations en provenance de son Environnement. En effet, on considère souvent la perception comme un phénomène Purement passif. Dans le monde de la robotique nous serions des robots mobiles munis de microphones (les oreilles), de caméras (les yeux), de nombreux capteurs de pression et de température (la peau), de capteurs chimiques (le goût et l'odorat) et de capteurs cinétiques (les labyrinthes ou vestibules : organes de l'oreille renseignant l'individu sur sa position dans l'espace et les mouvements qu'il y effectue). Dans notre cas, nous désirons développer la perception auditive. Donc, avant de commencer l'étude du capteur pour la localisation 2D, il m'a paru important de rappeler quelques définitions générales.

### 2 Généralités sur les signaux sonores

La perception naturelle de l'espace sonore passe par la localisation des sources qui nous Entourent. Localiser une source, c'est déterminer son azimuth (sa direction), sa profondeur (sa distance) et son élévation (sa hauteur). Un signal est un phénomène physique transportant une information. Nous nous intéressons ici aux signaux sonores qui peuvent être, par exemple, un 'clap', un sifflement, ou tout simplement un bruit. On sait qu'un signal sonore est une vibration de l'air se transmettant par des variations de pressions plus ou moins rapides et plus ou moins importantes, la rapidité étant caractérisée par la fréquence tandis que

l'importance est caractérisée par l'amplitude de la vibration, la puissance du signal est proportionnelle au carré de son amplitude. Les sons peuvent se propager dans tous les milieux matériels, mais, pour des raisons techniques, on ne peut les modifier qu'avec des appareils électroniques de sorte que, pour un son déterminé, il faut le transformer en signal électromagnétique pour lui faire subir le traitement que l'on désire. Tout comme la lumière, le son est constitué par un spectre. L'oreille est équipée pour l'analyse de ce spectre: elle est constituée de milliers de cellules spécialisées dans une gamme très sélective de fréquences. La perception est un procédé d'analyse du spectre sonore par décomposition en sons élémentaires. Notre oreille est sensible aux fréquences entre 20 Hz et 20kHz, avec un maximum de sensibilité aux alentours de 3 kHz. C'est aux alentours de cette fréquence que sont situés les sons produits par la voix. La fréquence d'un signal sonore caractérise la perception de hauteur de ce son: plus la fréquence est grande, plus le son perçu est aigu et inversement, plus la fréquence est basse, plus le son perçu est grave. L'amplitude sonore correspond à l'intensité, elle est mesurée en décibels acoustiques (dBa). 0 db correspond à notre seuil d'audition. Le carré de l'amplitude du signal est proportionnel à la sensation d'intensité sonore.

### **3 Etat de l'art des microphones**

Le premier axe de notre recherche consistait à étudier les différentes technologies pour la mesure d'un son. Il existe trois technologies majeures:



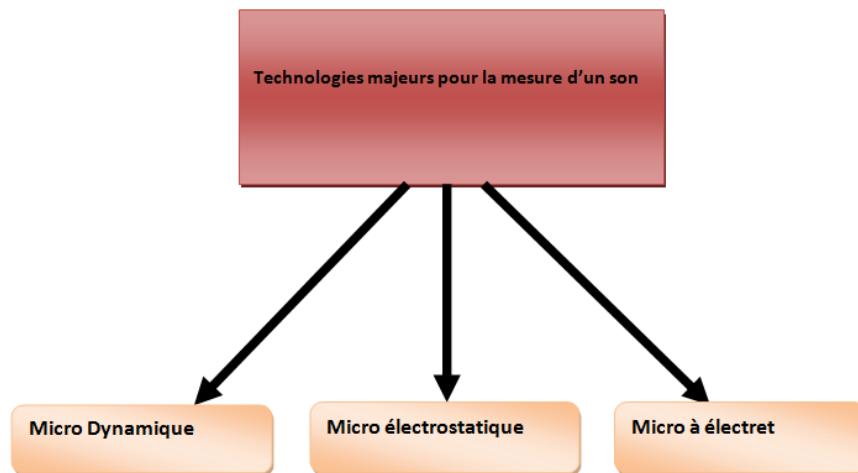


Figure: Représentation de technologies majeurs pour la mesure d' un son.

**a. Micro Dynamique**

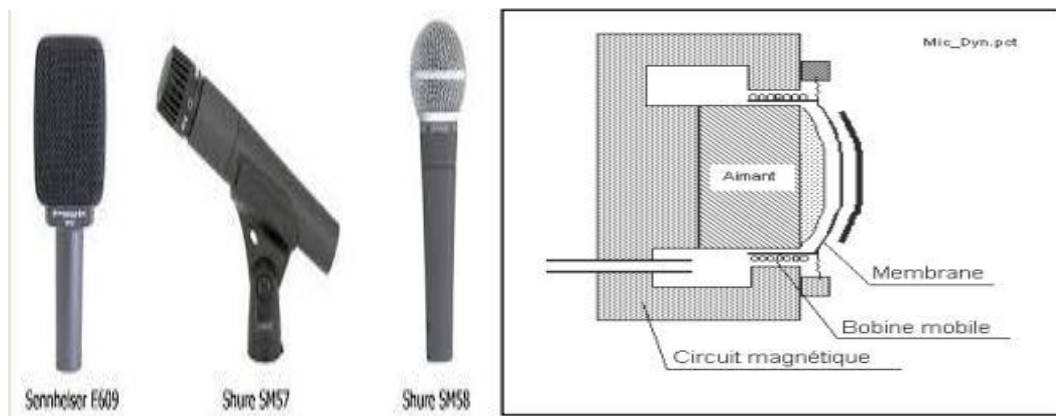


Figure: Fonctionnement d' un micro dynamique.

**b. Micro électrostatique**

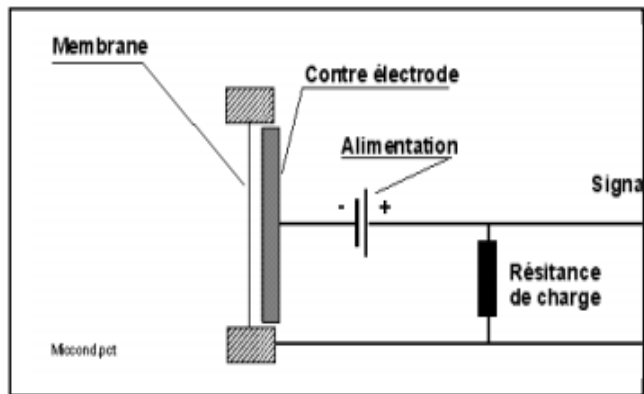


Figure : Fonctionnement d' un micro électrostatique.

**c. Micro à électret**

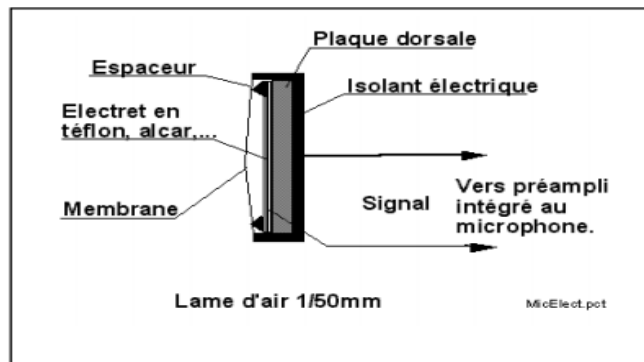
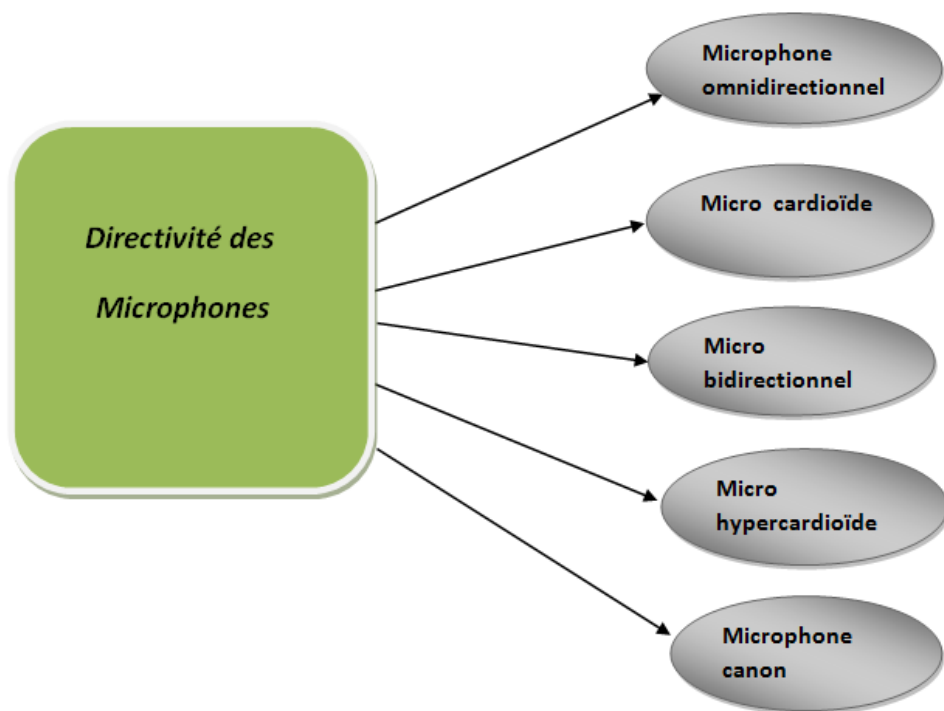


Figure: Fonctionnement de micro à électret.



**Figure:** Schéma de Directivité des Microphones.

## Bibliographie

---

- [1] [http://raphael.isdant.free.fr/traitement\\_numerique/3traitement\\_numerique\\_du\\_son](http://raphael.isdant.free.fr/traitement_numerique/3traitement_numerique_du_son).
- [2] Roger REYNAUD, "Cours de Traitement du Signal", Paris, p.3, (2003).
- [3] Didier Pietquin, "techniques du son/Sources sonores – Propagation", (2007).
- [4] <http://www.cochlea.eu/son/psychoacoustique/localisation>.
- [5] KHADDOUR, Hasen. "A comparison of algorithms of sound source localization based on time delay estimation". *elektrorevue*, p.31-37, (2011).
- [6] Haljaoui Adel, Brulin Damien, Val Thierry et Nasri Nejah, "Localisation d'une source sonore par un réseau de microphones", dans: Journées Nationales des Communications Terrestres - JNCT 2014, 22 Mai 2014 - 23 Mai 2014 (ToulouseFrance), (2014).
- [7] Murray, John C., Harry Erwin, and Stefan Wermter. "Robotics sound-source localization and tracking using Interaural time difference and cross-correlation". *AI Workshop on NeuroBotics*, (2004).
- [8] Cobos, M.Lopez, J.J. et Martinez, D., "Two-Microphone Multi-speaker Localisation based on a laplacian Mixture Model", *Digital signal Processing (Elsevier)*, Vol21, Issue 1, p. 66-67, (2011).
- [9] Yermehé, Z., "Soft-Constrained Subband Beamforming for speech Enhancement", these de doctorat, Sweden, (2007).

- [10] Varma,K.,“Time-Delay-Estimate Based Direction-of-Arrival Estimation for Speech in Reverberant Environment “ , these de Magister ,(2002).
- [11] MA, Hua. Localisation de sources par méthodes à haute résolution et par analyses parcimonieuse. Thèse de doctorat. (2011).
- [12] Gulay Yilmaz, Sound Source Localization by Using Two Microphones, p.3-7, (2014).
- [13] SRISHESHUKUMAR.B , “Microphone Array winner Beamforming with modeling of SRP-PHAT for Speaker Localisation ”, Master of science in Electrical Engineering with Emphasis of signal processing , p.28-31, ( January 2012).
- [14] Valin, J.M, Michaud.F , Rouat. J. et Létourneau, D. , “Robust Sound source Localization Using a Microphone Array on a Mobile Robot” , Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems , p.1228-1233.
- [15] Litva, J. et Kwok-Yeng lo, T., “Digital beamforming in Wireless Communications”, Artech House, Norwood, p.297, (1996).
- [16] Huang, Y. Benesty , J. et Chen, J., “Acoustic MIMO Signal Processing Handbook”, Springer-Verlag, berlin, p.391, (2006).
- [17] Madisetti , V. Et Williams, D.B., “Digital signal Processing Handbook”, CRC/IEEE Press , p.1690, (1999).
- [18] Van Veen , B.D, et Buckley , K.M., “Beamforming : A versatile Approach to Spatial Filtering”, IEEE Acoustic , Speech , and signal Processing(ASSP) Magazine , p.4-24, (April 1988).
- [19] SRISHESHUKUMAR.B , “Microphone Array winner Beamforming with modeling of SRP-PHAT for Speaker Localisation ”, Master of science in Electrical Engineering with Emphasis of signal processing , p.31-33, ( January 2012).
- [20] MA, Hua. Localisation de sources par méthodes à haute résolution et par analyses parcimonieuse. Thèse de doctorat, p.60, (2011).

[21] Grondin, François, and James Glass, " Multiple sound source localization with SVD-PHAT". arXiv preprint: [1906.11913](https://arxiv.org/abs/1906.11913) (2019).

[22] <http://iut-gmp.univ-lille1.fr/fichiers/LPVI/RapportFinalkinect>.

[23] Matthias, Kronlachner, "The Kinect distance sensor as human-machine-interface in audio-visual art projects", (2013).

[24] <https://www.alloprof.qc.ca/fr/elevés/bv/mathematiques/l-hyperbole-conique-m1329>.

[25] DOUAER, Belgacem, Farid YKHLEF, and Faycal YKHLEF." Tracking a sound Source in two Dimensions Using Microsoft Kinect Sensors". 2019 International Conference on Advanced Electrical Engineering (ICAEE). IEEE,(2019).